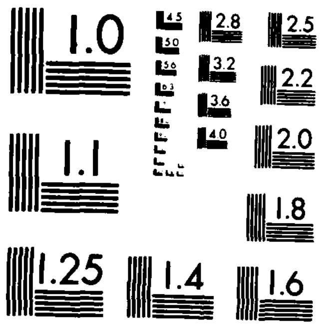
AD-A150 505 THE ORGANIZATION OF KNOWLEDGE IN A MULTI-LINGUAL INTEGRATED PARSER(U) YALE UNIV NEW HAVEN CT DEPT OF COMPUTER SCIENCE S L LYTINEN NOV 84 YALEU/CSD/RR-340 1/3

UNCLASSIFIED N00014-82-K-0149 F/G 5/7 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

The Organization of Knowledge In a Multi-lingual, Integrated Parser

Steven Leo Lytinen

YALEU/CSD/RR #340

November 1984

YALE UNIVERSITY
DEPARTMENT OF COMPUTER SCIENCE

This work was presented to the Graduate School of Yale University in candidacy for the degree of Doctor of Philosophy.

The Organization of Knowledge In a Multi-lingual, Integrated Parser

Steven Leo Lytinen

YALEU/CSD/RR #340
November 1984

This work was supported in part by the National Science Foundation under contract #IST-8017790, and by the Advanced Research Projects Agency of the Department of Defense, monitored by the Office of Naval Research under contract #N00014-82-K-0149.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 340

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle): The Organization of Knowledge In a Multi-lingual, Integrated Parser

5. TYPE OF REPORT & PERIOD COVERED: Research Report

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s): Steven Leo Lytinen

8. CONTRACT OR GRANT NUMBER(s): N00014-82-K-0149

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Yale University - Dept. of Computer Science
10 Hillhouse Avenue
New Haven, CT 06520

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS:
Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, Virginia 22209

12. REPORT DATE: November 1984

13. NUMBER OF PAGES: 244

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):
Office of Naval Research
Information Systems Program
Arlington, Virginia 22217

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report):
Distribution of this report is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Machine Translation
Natural Language
Parsing
Disambiguation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A controversy has existed over the interaction of syntax and semantics in natural language understanding systems. On the one hand, theories of integrated parsing have argued that syntactic and semantic processing must take place at the same time. In addition, these theories have also argued that syntactic and semantic knowledge should be mixed together, and that the role of syntax should be completely subserviant to semantic processing. On the other hand, opponents of this theory argue that parsing should be more modular, with syntactic and semantic processing taking place separately.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Along with this processing modularity, these opponents also argue that syntactic and semantic knowledge should be more modular, and that syntax, since it is largely autonomous from semantics, plays a more important role in natural language understanding.

This thesis presents a theory of natural language understanding which is a compromise between these two views. I argue that natural language understanding should be integrated, in the sense that syntactic and semantic processing should take place at the same time. However, instead of mixing syntactic and semantic knowledge together in the knowledge base of a parser, I argue that power can be gained by organizing syntax and semantics as two largely separated bodies of knowledge, which are combined only at the time of processing. The result is a parser which retains the predictive power which is gained by using semantic information during syntactic processing, but which is more robust in parsing complex syntactic constructions, and which is more amenable to the organization of knowledge about more than one language.

OFFICIAL DISTIRUBTION LIST

Defense Documentation Center Cameron Station Alexandria, Virginia 22314	12 copies
Office of Naval Research Information Systems Program Code 437 Arlington, Virginia 22217	2 copies
Dr. Judith Daly Advanced Research Projects Agency Cybernetics Technology Office 1400 Wilson Boulevard Arlington, Virginia 22209	3 copies
Office of Naval Research Branch Office - Boston 495 Summer Street Boston, Massachusetts 02210	1 copy
Office of Naval Research Branch Office - Chicago 536 South Clark Street Chicago, Illinois 60615	1 copy
Office of Naval Research Branch Office - Pasadena 1030 East Green Street Pasadena, California 91106	1 copy
Mr. Steven Wong New York Area Office 715 Broadway - 5th Floor New York, New York 10003	1 copy
Naval Research Laboratory Technical Information Division Code 2627 Washington, D.C. 20375	6 copies
Dr. A.L. Slafkosky Commandant of the Marine Corps Code RD-1 Washington, D.C. 20380	1 copy
Office of Naval Research Code 455 Arlington, Virginia 22217	1 copy

Office of Naval Research Code 458 Arlington, Virginia 22217	1 copy
Naval Electronics Laboratory Center Advanced Software Technology Division Code 5200 San Diego, California 92152	1 copy
Mr. E.H. Gleissner Naval Ship Research and Development Computation and Mathematics Department Bethesda, Maryland 20084	1 copy
Captain Grace M. Hopper, USNR Naval Data Automation Command, Code OOH Washington Navy Yard Washington, D.C. 20374	1 copy
Dr. Robert Engelmore Advanced Research Project Agency Information Processing Techniques 1400 Wilson Boulevard Arlington, Virginia 22209	2 copies
Professor Omar Wing Columbia University in the City of New York Department of Electrical Engineering and Computer Science New York, New York 10027	1 copy
Office of Naval Research Assistant Chief for Technology Code 200 Arlington, Virginia 22217	1 copy
Computer Systems Management, Inc. 1300 Wilson Boulevard, Suite 102 Arlington, Virginia 22209	5 copies
Ms. Robin Dillard Naval Ocean Systems Center C2 Information Processing Branch (Code 8242) 271 Catalina Boulevard San Diego, California 92152	1 copy
Dr. William Woods BBN 50 Moulton Street Cambridge, MA 02138	1 copy

Professor Van Dam
Dept. of Computer Science
Brown University
Providence, RI 02912 — 1 copy

Professor Eugene Charniak
Dept. of Computer Science
Brown University
Providence, RI 02912 — 1 copy

Professor Robert Wilensky
Univ. of California
Elec. Engr. and Computer Science
Berkeley, CA 94707 — 1 copy

Professor Allen Newell
Dept. of Computer Science
Carnegie-Mellon University
Schenley Park
Pittsburgh, PA 15213 — 1 copy

Professor David Waltz
Univ. of Ill at Urbana-Champaign
Coordinated Science Lab
Urbana, IL 61801 — 1 copy

Professor Patrick Winston
MIT
545 Technology Square
Cambridge, MA 02139 — 1 copy

Professor Marvin Minsky
MIT
545 Technology Square
Cambridge, MA 02139 — 1 copy

Professor Negroponte
MIT
545 Technology Square
Cambridge, MA 02139 — 1 copy

Professor Jerome Feldman
Univ. of Rochester
Dept. of Computer Science
Rochester, NY 14627 — 1 copy

Dr. Nils Nilsson
Stanford Research Institute
Menlo Park, CA 94025 — 1 copy

Dr. Alan Meyrowitz 1 copy
Office of Naval Research
Code 437
800 N. Quincy Street
Arlington, VA 22217

Dr. Edward Shortliffe 1 copy
Stanford University
MYCIN Project TC-117
Stanford Univ. Medical Center
Stanford, CA 94305

Dr. Douglas Lenat 1 copy
Stanford University
Computer Science Department
Stanford, CA 94305

Dr. M.C. Harrison 1 copy
Courant Institute Mathematical Science
New York University
New York, NY 10012

Dr. Morgan 1 copy
University of Pennsylvania
Dept. of Computer Science & Info. Sci.
Philadelphia, PA 19104

Mr. Fred M. Griffee 1 copy
Technical Advisor C3 Division
Marine Corps Development
and Education Command
Quantico, VA 22134

Acknowledgements

I want to thank my thesis advisor, Professor Roger Schank, for the role that he has played in advising my dissertation research, for his many helpful comments on the large number of drafts of this dissertation which he had the patience to read, and for the instrumental role that he has played in teaching me much of what I know about doing AI research.

I also want to thank the other members of my reading committee, Professor Drew McDermott and Dr. Christopher Riesbeck, who also read many drafts of the dissertation, and who provided valuable comments.

Other people read various versions of some or all of this thesis, or other papers which later turned into thesis chapters. They include Gregg Collins, Lawrence Birnbaum, Colleen Seifert, William Bain, Kristian Hammond, and Charles Martin. These people all provided me with helpful comments.

Many people contributed to the implementation of the MOPTRANS system, in its various incarnations. In the version presented in the dissertation, Eduard Hovy did much of the work on German generation, Thomas Grossi did all of the work on French parsing, and Wei-tung Li worked on Chinese parsing. Many people provided much assistance on previous versions of the system, including Professor Shoshana Hardt, Dr. Laurence Danlos, Dr. Shun Ishizaki, Abraham Gutman, Eduard Hovy, Thomas Grossi, David Littleboy, and Rod McGuire.

The MOPTRANS parser's skills in many languages far exceed my own. Thus, I required a great deal of language assistance in my work. I want to thank Abraham Gutman, Dr. Sylvia Galambos, Wei-tung Li, Eduard Hovy, W. Lewis Johnson, Dr. Laurence Danlos, David Littleboy, Professor Shoshana Hardt, Thomas Grossi, and William Bain for helping me with their expertise in the various languages in the MOPTRANS system, and for providing insights about other languages, particularly Japanese and Hebrew. The large number of languages which the MOPTRANS group discussed forced me to be more honest than I sometimes wanted to be in designing a multi-lingual parser.

Many people not directly associated with the MOPTRANS project have nevertheless had a large effect on the project. I would like to thank Professor Michael Lebowitz, under whom I served an apprenticeship, and whose project certainly influenced my own. Lawrence Birnbaum, Dr. Christopher Riesbeck, and William Ferguson also have had a large influence on the project through my discussions with them about parsing.

In addition, I would like to thank many other people in the Yale A.I. lab whose presence has inspired the research atmosphere in the lab, including Professor Elliot Solloway, Professor Janet Kolodner, Mark Burstein, Natalie Dehn, Professor Earnest Davis, Dr. Brian Reiser, Professor Scott Robertson, Stephen Slade, Peter Johnson, Stephen Salzberg, David Atkinson, David Miller, Valerie Abbott, Professor Robert Abelson, Dr Beth Adelson, Thomas Dean, Eric Gold, Yoav Shoham, James Spoher, and Steven Hanks.

Three people in particular are responsible for making me just sit down and work on my thesis even when I didn't want to. They are Colleen Seifert, Kristian Hammond, and Gregg Collins.

William Bain provided me with a great deal of support during the last few tense months when I never thought I would finish. I owe him a great deal of thanks.

Finally, the research presented in this dissertation was supported in part by the National Science Foundation under contract #IST-8017790, and by the Advanced Research Projects Agency of the Department of Defense, monitored by the Office of Naval Research under contract #N00014-82K-0149.

ABSTRACT

The Organization of Knowledge in a Multi-lingual, Integrated Parser

Steven Leo Lytinen

Yale University, 1984

A controversy has existed over the interaction of syntax and semantics in natural language understanding systems. On the one hand, theories of *integrated parsing* have argued that syntactic and semantic processing must take place at the same time. In addition, these theories have also argued that syntactic and semantic knowledge should be mixed together, and that the role of syntax should be completely subservient to semantic processing. On the other hand, opponents of this theory argue that parsing should be more modular, with syntactic and semantic processing taking place separately. Along with this processing modularity, these opponents also argue that syntactic and semantic knowledge should be more modular, and that syntax, since it is largely autonomous from semantics, plays a more important role in natural language understanding.

This thesis presents a theory of natural language understanding which is a compromise between these two views. I argue that natural language understanding should be integrated, in the sense that syntactic and semantic processing should take place at the same time. However, instead of mixing syntactic and semantic knowledge together in the knowledge base of a parser, I argue that power can be gained by organizing syntax and semantics as two largely separate bodies of knowledge, which are combined only at the time of processing. The result is a parser which retains the predictive power which is gained by using semantic information during syntactic processing, but which is more robust in parsing complex syntactic constructions, and which is more amenable to the organization of knowledge about more than one language.

Table of Contents

List of Figures

1. Introduction

1.1 What is Integrated Parsing?

Consider the following sentences:

John made a reservation for two people on the 9:30 flight to Los Angeles.

John made the slides for his presentation on the 9:30 flight to Los Angeles.

The travel agent made a rent-a-car reservation for two people on the 9:30 flight to Los Angeles.

Syntactically, these three sentences are all ambiguous. In each sentence, the prepositional phrase "on the 9:30 flight to Los Angeles" can be attached to one of three places: to the verb "made," to the direct object of "made" ("reservation" or "slides"), or to the object of the preposition "for." However, despite these syntactic ambiguities, a human reader understands, unambiguously, the meaning of all three of these sentences. This is because the contexts in which "on the 9:30 flight to Los Angeles" appears above provides enough information to determine which attachment makes the most sense.

These examples illustrate that decisions about the syntactic structure of a sentence must sometimes be influenced by *semantic knowledge*, or knowledge about the meanings of words; and *pragmatic knowledge*, or knowledge about the world and about how language is used. In order to determine where to attach the prepositional phrase in these three sentences, one must know that it is possible to make a reservation for the 9:30 flight to Los Angeles, but not a rent-a-car reservation for this flight. One must also know that presentations in which slides are shown are not typically done on an airplane, but that the slides could be prepared on an airplane. All of this semantic/pragmatic knowledge must be used to eliminate the syntactic ambiguities in the three examples.

Examples like these support the argument for an *integrated* approach to natural language analysis. In this approach, morphological, syntactic, semantic, and pragmatic processing are all performed at once, so that all types of information are available to any parsing decisions that are made at any of these levels. This seems necessary for the above examples, because of the semantic/pragmatic knowledge that needs to be referenced in order to make the correct syntactic decisions. A parser which makes its syntactic decisions without complete access to this knowledge would make mistakes in these examples, or at least finish its parse of these sentences with an ambiguity remaining.

It seems difficult to limit in any way the type of semantic/pragmatic knowledge that might be needed to make syntactic parsing decisions. In the examples above, the necessary inferences are not trivial. For example, in order to conclude that "on the 9:30 flight to Los Angeles" should not be attached to "presentation" in the second example, we must use a great deal of world knowledge about what type of presentation involves the use of slides. It is not enough simply to know that presentations are not usually done on airplanes, as the following example illustrates:

I wrote United airlines to complain about the movie presentation on the 9:30 flight to Los Angeles.

In this example, because we know that movies are often shown on airplanes, it makes sense to attach "on the 9:30 flight to Los Angeles" to "presentation."

Thus, we need the full power of semantic/pragmatic processing to make syntactic decisions like these. We need the integration of processing to be complete; that is, no portion of semantic/pragmatic processing can be separated from or postponed until after morphological/syntactic processing.

1.2 Integrated Parsing and the Syntax/Semantics Controversy

A battle has raged in natural language processing for many years over how syntax and semantics[1] should interact with each other. There are many different dimensions along which this question can be asked, such as the following:

- What is the order in which syntactic and semantic processing take place during the understanding of a text?
- How much interaction is there between syntactic and semantic processing?
- How important are the roles that syntax and semantics play in the process of understanding the meaning of a text?
- How should syntactic and semantic knowledge be represented; i.e., should there be separate bodies of syntactic and semantic knowledge, or should they be mixed in some way?

Any theory of integrated parsing, as I defined it in section 1.1, must take a stand on the first and second of these questions: it must assert that syntactic and semantic processing should take place at the same time, and that a great deal of interaction between these types of processing is necessary. However, believing in integrated parsing does not necessarily entail a particular belief with regards to the last two questions. Syntactic processing could conceivably play a very important or very unimportant role in an integrated parser; likewise, although syntax and semantics are *processed* together, syntactic and semantic *knowledge* could be stored completely separately, or could be completely mixed together, or somewhere in between. Integrated processing claims say nothing about representation of knowledge.

However, previous advocates of integrated parsing have, for the most part, taken stands on these two issues. Moreover, these stands have been in direct opposition to those who advocate non-integrated parsing. Thus, the two sides have lined up, opposed to each other on all of the issues concerning the interaction of syntax and semantics that I have mentioned. The views of two sides are as follows:

Proponents of integrated parsing:[2] Syntax plays a relatively unimportant role in the process of understanding natural language. Semantics guides the parsing process, and calls on syntax only when it needs to. Syntactic

---

[1]From here on, for the sake of brevity, I will simply use *semantics* or *semantic knowledge* to refer to both semantics and pragmatics.

[2]Arguments for this view can be found in (Wilks, 1975a), (Riesbeck and Schank, 1976), (Small, 1980), (Schank and Birnbaum, 1980), (Lebowitz, 1980).

and semantic processing proceed at the same time, with no separate syntactic representation of a text necessary. Communication between syntax and semantics is high. Syntactic decisions are made with full access to all semantic processing which has been performed. Knowledge about syntax and semantics is highly mixed (although there may be some purely syntactic knowledge), with syntactic knowledge *encoded in a largely procedural form, often referring to semantics.*

Opponents of integrated parsing:[3] Syntax plays an important role in the process of understanding the meaning of a natural language text. Syntactic and semantic processing are largely separate, with syntactic processing performed first (although semantic processing can be interleaved with syntactic processing; i.e., once syntax has produced a partial analysis, semantic interpretation of that portion of the text can proceed before other portions of the text are syntactically analyzed). Syntax and semantics interact with each other in limited ways, if at all. Syntax might be allowed to ask certain types of questions of semantics at particular times, but communication between syntactic and semantic processing is not unlimited. Knowledge about syntax and semantics is also largely separate. Syntactic knowledge can be expressed without much reference to semantics.

Examples of parsers written by proponents of integrated parsing include Wilks' parser (Wilks, 1973) (Wilks, 1975a), ELI (Riesbeck, 1975), the Integrated Partial Parser (IPP) (Lebowitz, 1980), the Word Expert Parser (Small, 1980), and BORIS (Dyer, 1982). In these parsers, there was no distinction between syntactic or semantic processing of a sentence. All different kinds of knowledge were available to the parsing process at all times. The result of this simultaneous application of knowledge was the immediate building of a representation of the meaning of the text, without the building of intermediate syntactic representations. The representational systems used in these parsers consisted of primitives such as Conceptual Dependency (Schank, 1972) or those used by Wilks (Wilks, 1973); frames (Minsky, 1975); or scripts (Schank and Abelson, 1977).

In these parsers, there was no distinction between syntactic and semantic rules. Syntactic and semantic knowledge was compiled together into their rule bases. For example, in BORIS, the following parsing rules were used to fill the slots of the verb "grading" in the sentence "John was grading homework assignments":

> If a HUMAN appears before the word "grading," then assign that HUMAN to be the EVALUATOR of the action GRADE.
>
> If a WORK-OBJ [a class of physical objects] appears after the word "grading," then assign the WORK-OBJ to be the OBJECT of the action GRADE.

These rules contain syntactic knowledge, that a noun group to the left of the word "grading" fills the EVALUATOR slot of the action GRADE, and a noun group to the right of "grading" fills the OBJECT slot. They also contain semantic/pragmatic knowledge, that the EVALUATOR of the action GRADE should be a HUMAN, and the OBJECT of GRADE should be a WORK-OBJ.

---

[3] Arguments for this view can be found in (Chomsky, 1965), (Woods, 1970), (Marcus, 1978), (Hirst, 1983).

Examples of the non-integrated approach to natural language processing include systems which use ATN parsers, such as LUNAR (Woods, Kaplan and Nash-Webber, 1972); PARSIFAL (Marcus, 1978); and Winograd's parser (Winograd, 1972). These parsers produced syntactic analyses of input texts, with limited reference to semantics or pragmatics. Then the results of the syntactic analysis were passed to a semantic interpretation phrase, which operated on the parse tree to extract whatever semantic information was required of the system (e.g., blocks-world operations in Winograd's parser).

The rule bases in these parsers consisted of largely separate bodies of syntactic and semantic knowledge. For example, Winograd's parser contained procedurally-encoded versions of phrase structure grammar rules such as the following:

```
S -> NP VP
NP -> DETERMINER NOUN
VP -> VERB/TRANSITIVE NP
VP -> VERB/INTRANSITIVE
```

Once rules like these produced a syntactic parse tree, separate semantic rules were applied to build the semantic representation, which was then used to manipulate the blocks world or to answer questions.

1.3 The Claims of This Thesis

The goal of this thesis is to show that the views held by both sides of the syntax/semantics controversy are too extreme. As an alternative, I will argue for a theory of natural language processing which entails some of the claims of both sides. The result is a parser which is integrated in the sense that I defined in section 1.1, but which uses syntax to a larger degree than previous integrated parsers, and has a largely separate body of syntactic knowledge.

This thesis discusses the interaction of syntax and semantics with respect to the task of *conceptual analysis*. By conceptual analysis, I mean the task of building a representation of the meaning of a text (as opposed to *syntactic analysis*, which is the task of building a representation of the syntactic structure of a text). In particular, I argue for the following claims with regards to the interaction between syntax and semantics in conceptual analysis:

1. **Syntactic and semantic processing of a text should proceed at the same time.**
2. **Syntactic decisions must be made with full access to semantic processing; that is, communication between syntax and semantics is high.**
3. **A limited amount of syntactic representation must be built during text understanding.**
4. **Knowledge about syntax and semantics is largely separate. Syntactic knowledge should be expressed in the parser's knowledge base as a largely separate body of knowledge, but this knowledge should have references to semantics, telling the system**

how semantic representations are built from these syntactic rules.

5. **Semantics guides the parsing process, but relies on syntactic rules to make sure that it is making the right decisions.**

To demonstrate the advantages of this theory of natural language processing, I will present an integrated, multi-lingual parser which parses short (1-3 sentences) newspaper articles about terrorism and crime in English, Spanish, French, German, and Chinese. This parser produces language-independent, conceptual representations for the stories that it reads, similar to the representations which previous integrated parsers have produced. It operates as part of a machine translation system, called MOPTRANS. Enough vocabulary, linguistic knowledge, and semantic knowledge have been encoded in the parser to enable it to parse 15-50 stories for each input language. The MOPTRANS system produces translations for all of the stories into English, and for some of the stories into German. The stories, the representations produced by the parser, and the English translations produced by the MOPTRANS system are found in appendix 1.

Because the MOPTRANS parser is integrated, in the sense that syntactic and semantic processing proceeds in parallel, syntactic decisions are made with full access to the results of semantic processing. Thus, unlike non-integrated parsers, MOPTRANS uses all of the semantic knowledge available to it to resolve syntactic ambiguities such as in the examples I presented in section 1.1. I will demonstrate MOPTRANS' advantages over non-integrated, syntactic parsers by discussing the difficulties that these parsers have in dealing with the problems of machine translation, and how MOPTRANS overcomes these problems.

Because of the modifications to previous theories of integrated parsing, the MOPTRANS parser also has advantages over previous integrated parsers, with respect to the following problems:

Frame Selection

One issue which has arisen in conceptual analysis is due to the use of frames (Minsky, 1975) and other frame-like structures such as scripts (Schank and Abelson, 1977) to represent the meaning of the text. The *frame selection problem* (Charniak, 1982), or the selection of the appropriate frame for a text, must be faced by any conceptual analyzer which knows a large number of possible frames. Sometimes, particular words in a text point directly to a particular frame, thus trivializing this problem. For example, the word "arrest" refers directly to a high-level structure, such as the $ARREST script. However, more often it is the case that no one word in a text points definitively to a unique frame. Instead, many of the words in the text are ambiguous or vague, and it is only by considering them in combination that a frame can be selected. An arrest, for instance, can be described without using the word "arrest," as in "Police took a suspect into custody," or even "They got their man." In cases like this, frame selection is much more difficult.

If syntactic and conceptual knowledge are not separated, I will show that solving the frame selection problem requires the use of an unmanageably large number of frame selection rules. However, with largely separate bodies of syntactic and conceptual knowledge, the MOPTRANS parser is able to perform frame selection for difficult examples encountered in its newspaper articles, involving very vague words or phrases, using only a few purely semantic concept refinement rules.

Parsing Complex Syntactic Constructions

Past integrated parsers have for the most part not attempted to parse syntactically complex texts, or else have settled for a partial parse or a skimming of these sentences (e.g., IPP (Lebowitz, 1980) and FRUMP (DeJong, 1979)). Some attempts have been made to parse constructions such as unmarked relative subclauses, but it is not clear how robust these attempts have been. For example, the following sentence was parsed by the Conceptual Analyzer (CA) (Birnbaum and Selfridge, 1979):

A small plane stuffed with 1500 pounds of marijuana crashed.

Birnbaum and Selfridge proposed the following parsing rule to identify the relative subclause:

Test: The word "with" follows the word "stuffed," followed by a noun group which could function semantically as the OBJECT of the action "stuffed."

Action: Fill the OBJECT of the action "stuffed" with the noun group following "with"; mark the noun group to the left of "stuffed" as the thing being stuffed.

By performing the slot-fillings in the ACTION portion of this rule, CA in effect recognized that "stuffed" was being used as an unmarked passive.

This rule relies on the appearance of a key preposition after the unmarked passive to identify that this is in fact the syntactic function of the past participle. In general, though, it is not clear that this approach would work, as the following example demonstrates:

The soldier called to his sergeant.
I saw the soldier called to his sergeant.

Here we see that the preposition "to" can appear after "called" whether called is active or passive. Thus, since "to" could not be used as a signal indicating an unmarked relative subclause, it is not clear how this approach would be able to handle examples like these.

With more autonomous syntactic knowledge, the MOPTRANS parser is able to reliably handle complex syntactic constructions in many different languages. The constructions include many types of marked and unmarked clauses, the use of present participles as nouns, infinitive phrases, and many cases of conjunction.

Multi-lingual Parsing

Writing a conceptual analyzer which can process inputs from more than one language requires a certain modularity of the knowledge used by the parser. Some parsing knowledge, or parsing rules, must be shared between languages; otherwise the parser is not multi-lingual in any interesting sense. If nothing is shared, then one might just as well write separate parsers for each language.

In a multi-lingual conceptual analyzer, much of the parser's semantic knowledge should be sharable among different languages. After all, the same knowledge about the world should be applicable to the processing of different languages. Since syntactic knowledge varies from language to language, though, syntactic knowledge about each

particular language must be stored separately from semantic knowledge in order to facilitate sharing. Thus, it would not be possible in previous integrated parsers to facilitate this sharing of knowledge across languages. However, in the MOPTRANS parser, much of the body of semantic knowledge is used to parse all of the languages in the system.

Learning

Although this thesis will not propose any theories of language learning, this is an important issue that must be addressed in natural language research. Theories of natural language processing ought to be compatible with the task of language learning; i.e., at least some of the parsing knowledge proposed in these theories should be learnable.

The representation of parsing knowledge used in previous integrated parsers does not lend itself well to the task of learning. This is because it is difficult to identify the scope of a piece of parsing knowledge, since this knowledge, whether syntactic or semantic, is interwoven with other types of knowledge. For example, the BORIS parsing rules above contain syntactic knowledge which is applicable to all verbs, that the noun group before the verb and the noun group after the verb have a particular semantic relationship with the verb. Usually the noun group before the verb functions as some sort of AGENT of the verb (in this case, the EVALUATOR of the action GRADE), and the noun group after the verb functions as some sort of PATIENT (the OBJECT slot, in this case). However, with the type of rules used in previous integrated parsers, such as those used in BORIS, the fact that this syntactic information is applicable to most verbs is not marked. Instead, every verb to which this syntactic information applies has a rule containing this information in some form. Thus, a learning system using this sort of rule base would not know what syntactic knowledge would be applicable to a newly-learned verb. This lack of knowledge about the scope of a rule poses problems for a learning system.

In the MOPTRANS parser, since conceptual and syntactic rules are more autonomous, they are expressed at a more general level. Thus, the scope of the parser's rules is easily determined. Although MOPTRANS is not a language learner, this organization of parsing knowledge is more well-suited to the task of learning.

1.4 The Machine Translation Problem

Non-integrated, or syntactic, parsing approaches have been used in the past in the task of machine translation (e.g., (MacDonald, 1963), (Slocum and Bennett, 1982), (Boitet and Nedobejkine, 1981)). However, there are problems with translating texts which contain lexical or structural ambiguities using this approach. Often, the resolution of these ambiguities is essential to the ability to produce good translations. For example, consider the following English newspaper story, and its translation to German:

English: Black nationalists claimed on Monday that they were responsible for the midnight bombings at two strategic government oil refineries that killed 2 men and set off the worst fire in South Africa's history.

German: Schwartze Nationalisten behaupteten am Montag dass sie verantwortlich waren fuer die mitternaechtlichen Bombenangriffe bei zwei strategischen Regierungsoelraffinaderien dass 2 Maenner toeteten

und die schlimmste Feuer in der Geschichte von Suedafrika verursachten.

Literally in English: Black nationalists claimed on Monday that they responsible were for the midnight bombings at two strategic government-oil-refineries that two men killed and the worst fire in the history of South Africa caused.

What are the difficulties in performing this translation by computer? First, there is the matter of deciding how to translate the ambiguous words in the source text. The word "fire" in English can refer to the burning of something, the shooting of a weapon, or the letting go of an employee. The corresponding word in the above German translation, "Feuer," is only appropriate for the first of these meanings. Similarly, the phrase "set off" has been translated in this example as "verursachten" (caused). This is not the only way that this phrase could be translated. For instance, "Terrorists set off a bomb" would be translated into German as "Terroristen entzuendeten eine Bombe."

There is also a structural ambiguity in this sentence, which affects the way that it should be translated. Knowing which verbs are conjoined by "and" in the input sentence is essential to knowing how the portion of the sentence after "and" should be translated. This is because in German, verbs which are inside of relative clauses come at the end of the clause. Thus, since "and" conjoins "killed" and "set off" in this example, "set off" is inside of a relative clause, and the German verb, "verursachten," comes at the end of the sentence. If "and" conjoined "set off" and "claimed," however, "verursachten" would not come at the end of the sentence, because it would not be part of a relative clause.

Performing a semantic analysis of the input sentence is essential to the ability to resolve the ambiguities in this example. In order to determine that "fire" should be translated as "Feuer," and "set off" as "verursachten," the system must know that the story is saying that an explosion caused a fire, and that it is possible for an explosion to cause a fire. Similarly, this knowledge must be used in order to determine that "and" conjoins "killed" and "set off" instead of "claimed" and "set off." To determine this, the system must build a conceptual representation of the text, and check to see if the representation it is building is a reasonable one, according to the world knowledge that it has.

1.5 MOPTRANS: A Semantics-based Approach to Machine Translation

MOPTRANS (MOP-based TRANSlator) is an attempt to address some of the problems of machine translation in a semantics-based way. MOPTRANS is divided into two parts: an integrated multi-lingual conceptual analyzer, and a conceptual generator which produces the translation from the representation produced by the parser. The generator will not be discussed in this thesis.

MOPTRANS' parser and generator share a common conceptual knowledge base. This knowledge base consists of "world knowledge" facts about the domain of terrorism and crime, including the different types of events which can take place within this domain, and the actors, physical objects, etc., which are likely to play a part in these events. The same knowledge is used in the parser to parse all of the input languages, and in the generator to produce both the English and German translations. Because of this,

MOPTRANS is a truly interlingual system.

Linguistic knowledge is not shared between the parser and generator. This is not meant to be a theoretical claim, however; the decision to use separate linguistic rules in the parser and generator was purely a pragmatic one. Just as it is desirable to share conceptual knowledge as much as possible between the parser and the generator, any linguistic knowledge that could be shared would be desirable, also.

```
                 ----- S1                        T1 -----
                 |                                      |
                 +----- S2                       T2 -----+
Linguistic       |                                      |   Linguistic
Knowledge        +----- S3                       T3 -----+   Knowledge
for              |                                      |   for
Source           |       .                       .      |   Target
Languages        |       .                       .      |   Languages
                 |       .                       .      |
                 |                                      |
                 +----- Sn                       Tn -----+
                 |                                      |
                 |           ----------------           |
              ----------     | Intermediate   |     -------------
              | Parser |-->  | Conceptual     | -->| Generator   |
              ----------     | Representation |     -------------
                      \      ----------------      /
                       \           /\             /
                        \          ||            /
                         \  ------------        /
                            | Conceptual |
                            | Knowledge  |
                            ------------
```

Figure 1-1: Structure of the MOPTRANS System

In the translation process, an input story is first fed to the parser. Depending on the language of the input story, the parser uses the appropriate package of linguistic knowledge for that language to produce a conceptual representation of the meaning of the story. This conceptual representation is meant to be language-independent. Thus, the same representational system is used for all languages, and for a given input story, the same representation is built, independent of the source or target language. Once built, the representation is passed to the generator, which produces the translation in the target language, using its linguistic knowledge for that particular language. Figure 1-1 illustrates the structure of the system.

1.6 An example from MOPTRANS

Here is a sample of the output of the MOPTRANS system:

```
MOPTRANS created 27-Oct-83 13:13:18, ready  5-Jun-84 13:43:10

*(PARSE EN18)

Story EN18:

English:
Black nationalists claimed on Monday that they were responsible for the
midnight bombings at two strategic government oil refineries that killed
two people and set off the worst fires in South Africa's history.

Final representation:

FIR0 =
 CONCEPT     FIRE
 DEGREE      WORST
 LEAD-FROM   EXP0 =
               CONCEPT EXPLODE-BOMB
               LEAD-TO FIR0
               RESULT  DEA0 =
                          CONCEPT DEAD
                          R1      HUM1 =
                                    CONCEPT PERSON
                                    GENDER  MALE
                                    NUMBER  2
               ACTOR   HUM0 =
                          CONCEPT TERRORIST
                          RACE    BLACK
               PLACE   LOC0 =
                          CONCEPT  BUILDING
                          OWNED-BY ORG0 =
                                     CONCEPT AUTH-ORG
                                     OWNS    LOC0
                          NUMBER   2
               TIME    INS1 =
                          CONCEPT     INSTANCE
                          TIME-OF-DAY MIDNIGHT
 DURING-TIME DUR0 =
               CONCEPT DURATION
               OF      LOC1 =
                          CONCEPT NATION
                          WORD    south-africa
                          #NAME   SOUTH-AFRICA
CLA0 =
 CONCEPT CLAIM
 OBJECT  ACT0 =
```

```
          CONCEPT ACTOR
          R1      EXPO
          R2      HUMO
          TIME    INSO =
                    CONCEPT INSTANCE
                    DAY     MONDAY
 ACTOR   HUMO
```

(GEN 18)

Translation into German:

Nationalisten behaupteten am Montag dass sie verantwortlich waren fuer die Bombenangriffe bei zwei Raffinaderien dass 2 Maenner toeteten. Die Bombenangriffe verursachten die schlimmste Feuer in der Geschichte von Suedafrika.
NIL

The representation produced by the parser is shown above. It is languag meaning that this same representation, or a very similar representation, woul produced for versions of this story in other languages. For example, here is MOPTR. output for the German version of this story:

MOPTRANS created 27-Oct-83 13:13:18, ready 5-Jun-84 21:19:43

*(PARSE G18)

Input story:

Schwartze Nationalisten behaupteten am Montag dass sie verantwortlich war fuer die mitternaechtlichen Bombenangriffe bei zwei strategischen Regierungsoelraffinaderien dass 2 Maenner toeteten und die schlimmste Feu in der Geschichte von Suedafrika verursachten.

Final representation:

```
FIR7 =
 CONCEPT    FIRE
 DEGREE     WORST
 LEAD-FROM  EXP5 =
              CONCEPT EXPLODE-BOMB
              LEAD-TO FIR7
              ACTOR   HUM22 =
                        CONCEPT TERRORIST
                        RACE    BLACK
              PLACE   LOC4 =
                        CONCEPT  BUILDING
                        OWNED-BY ORG6 =
                                   CONCEPT AUTH-ORG
                                   OWNS    LOC4
```

```
                 NUMBER   2
        TIME    INS4 =
                 CONCEPT     INSTANCE
                 TIME-OF-DAY MIDNIGHT
 DURING-TIME DUR1 =
        CONCEPT DURATION
        OF      LOC5 =
                 CONCEPT NATION
                 #NAME   SOUTH-AFRICA
CLA0 =
 CONCEPT CLAIM
 OBJECT  ACT0 =
          CONCEPT ACTOR
          R1      EXP5
          R2      HUM22
 ACTOR   HUM22
 TIME    INS3 =
          CONCEPT INSTANCE
          DAY     MONDAY

Total time:  90400 msecs.
NIL

(GEN 18)

Translation into English:

Black nationalists claimed responsibility on Monday for bombings at midnight
in 2 refineries owned by the government that set off the worst fires during
the history of South Africa.
```

As I will argue in this thesis, a semantics-based approach to machine translation is important because of the difficulties of translating vague or ambiguous words. To demonstrate this, I will discuss the following example throughout the thesis:

```
MOPTRANS created 27-Oct-83 13:13:18, ready 5-Jun-84 21:19:43

*(PARSE SP6)

Input story:

 La policia realiza intensas diligencias para capturar a un presunto
 maniatico sexual que dio muerte a golpes y a punaladas a una mujer
 de 55 anos, informaron fuentes allegadas a la investigacion.

Literally in English:

 The police are realizing intense diligent actions in order to capture
 a presumed sex maniac who gave death by hits and stabs to a woman
 of 55 years, informed sources close to the investigation.
```

Final representation:

```
MTRO =
 CONCEPT MTRANS
 ACTOR   HUM5 =
           CONCEPT PERSON
 OBJECT  HARO =
           CONCEPT HARM-PERSON
           INST    OBJ1 =
                     CONCEPT WEAPON
                     INST-OF HARO
           ACTOR   HUM2 =
                     CONCEPT BAD-GUY
                     TYPE    SEX-MANIAC
                     STATUS  PRESUMED
           OBJECT  HUM4 =
                     CONCEPT PERSON
                     GENDER  FEMALE
                     AGE     YEA0 =
                               CONCEPT YEAR
                               NUMBER  55
           RESULT  DEA0 =
                     CONCEPT   DEAD
                     R1        HUM4
                     RESULT-OF HARO
*DO0 =
 CONCEPT POLICE-INVESTIGATION
 OBJECT  HUM2
 GOAL    GET0 =
           CONCEPT   ARREST
           GOAL-OF   *DO0
           ACTOR     HUM0 =
                       CONCEPT AUTHORITY
                       ORG     ORG0 =
                                 CONCEPT AUTH-ORG
                                 MEMBERS HUM0
           OBJECT    HUM2
 ACTOR   HUM0
 DEGREE  INTENSE

Total time:  238984 msecs.
NIL

(GEN 18)
```

Translation into English:

The police are searching for a presumed sex maniac who beat a 55-year-old woman to death.

The Spanish phrase "realizar diligencias" is very vague, and cannot easily be translated directly into English. Literally, the phrase means "to realize diligent actions." Often, however, the best translation of the phrase is dictated by its surrounding context, which can provide clues as to what specific action this vague phrase refers to. In this case, since the police are performing the diligent action, and the goal of the action is to capture a criminal, we can infer that the action is an investigation. Thus, a good translation for the phrase in this sentence is to use the English verb "to investigate." The inference abilities of the MOPTRANS parser allow it to come to the same conclusion, thus producing the above translation.

1.7 An Overview of the Thesis

The rest of this thesis will be devoted to discussing the theory of integrated natural language processing which I outlined in section 1.3, and the application of this theory to the task of machine translation. Chapter 2 will discuss at length syntactic approaches to machine translation, and their limitations. In chapter 3, I will explore the reasons why conceptual analysis can be of help to these problems.

Having motivated the use of conceptual analysis in machine translation, the remainder of the thesis will be devoted primarily to the discussion of multi-lingual conceptual analysis. In chapter 4, I will discuss previous research in integrated parsing, and explain why the integration of knowledge in these parsers has caused them to fall short in the solution of the problems which I discussed above. In chapters 5 and 6, I will present the approach to the representation of semantic and syntactic knowledge in the MOPTRANS parser, and why this approach does not suffer from the limitations of the integrated knowledge approach used in previous integrated parsers.

Finally, in chapter 7, I will compare the parsing rules used by MOPTRANS for the five different languages which it parses. Some linguistic phenomena, such as conjunction and pronominal reference, are handled in all languages by the same parsing rules. Other syntactic constructions are handled by different variations, depending on the language.

2. A Critique of Syntactic Machine Translation Methods

2.1 Introduction

Research in machine translation has generally focused on the translation of text by means of syntactic methods of analysis. When machine translation research first began in the late 1940's and the 1950's, researchers were optimistic that syntactic methods would result in the development of FAHQT (Fully Automated High-Quality Translation) in the not too distant future[4]

However, in the late 1950's and early 1960's, it became evident that the achievement of FAHQT was not very close. (Bar-Hillel, 1960) argued that fully automatic high-quality machine translation (FAHQT) was not feasible using the syntactic and table look-up techniques of his time. He argued that it was not always possible to correctly translate even very simple sentences without an "encyclopedia of knowledge" to refer to. An example he gave was "The box is in the pen," where the word "pen" should be translated as meaning "playpen," instead of a writing implement. In certain contexts, this would be the natural translation of this word:

> Little John was looking for his toy box. Finally he found it. The box was in the pen. John was very happy.

Bar-Hillel argued that the only way to determine that "pen" means "playpen" in this context would be to refer to knowledge about the relative sizes of writing implements, toy boxes, and playpens. This knowledge would provide the informatin that the referent of "pen" in this case must be a playpen. Since this sort of encyclopedic knowledge was not even proposed to be used in MT systems of that time, Bar-Hillel concluded that FAHQT was not a feasible goal.

During the 60's, machine translation research started a decline. In 1966, the Automatic Language Processing Advisory Committee (ALPAC) also concluded that FAHQT was not something soon to be achieved (see (ALPAC, 1966)), and funding for machine translation research in the United States was drastically reduced.

Now, however, MT research has resumed, in Europe, Canada, and even the United States. Most of this research, with the exception of Wilks' (1973) system, remains heavily syntax-based. It has changed in two ways from the earlier MT research. First, the goal of this research has generally been reduced from FAHQT to semi-automatic translation, in which the MT system produces an output which requires postediting by a human translator in order to produce the final translation. Second, limited amounts of semantic information have been introduced into many syntactic MT systems in order to try to solve some of the problems with earlier syntax-based MT research, and to hopefully reduce the amount of postediting necessary. With these changes it is now hoped that

[4]According to (Bar-Hillel, 1960), early progress in MT "created among many of the workers actively engaged in this field the strong feeling that a working system is just around the corner."

syntax-based MT[5] research will now be more successful.

This hope is encouraged by the fact that there are existing syntax-based systems already in use in limited commercial applications. Two such systems are TAUM-METEO (Chandioux, 1976), which has been in operation since 1976, translating Canadian weather reports between English and French; and SYSTRAN (Toma, 1977), which is in limited use by the European Economic Community. Currently, however, these systems are quite restricted, in that they only operate in severely restricted domains (such as TAUM-METEO), or require large amounts of human postediting (such as SYSTRAN).

In this chapter, I will examine the current syntactic MT research and its prospects for success. Although limited successes have already been achieved, I will argue that given the small amount of semantics in present-day syntactic MT systems, it is still not possible for syntactic MT research to attain FAHQT or even semi-automated translation with only small amounts of human postediting. The limited amount of semantic information used in present-day syntactic MT systems is inadequate for handling the problems involved with resolving ambiguities, and therefore any MT system which is primarily based on syntactic analysis techniques must make many errors when dealing with texts which contain ambiguous words or ambiguous syntactic structures, resulting in the need for a great deal of postediting. Only in a severely limited domain or with a syntactically limited subset of a natural language can syntax-based machine translation achieve the results of FAHQT or semi-automated translation with small amounts of postediting.

2.2 The General Syntactic Approach to Machine Translation

2.2.1 The Phases of Syntactic MT

Syntactic machine translation tends to be a 3-phase process. First, an *analysis phase* produces a syntactic parse tree from an input text in the source language. Second, a *transfer phase* transforms the parse tree from the analysis phase into a tree which is more appropriate for the target language. This involves substituting lexical items from the target language for the source language lexical items in the original parse tree (*lexical transfer*), as well as transforming the structure of the parse tree, in case the target language does not use an equivalent syntactic construction as was used in the source language (*structural transfer*). Finally, a *generation phase* produces the appropriate text in the target language from this transformed parse tree.

To illustrate the division of labor into these three phases, consider the following translation:

English: The fish like to swim.

German: Die Fische schwimmen gern.

[5] Despite the limited amounts of semantics that have been introduced to these systems, I will continue to refer to them as syntax-based so as to contrast them to the much more heavily semantics-based methods which I will discuss in later chapters.

```
                S
             /     \
           NP        VP
           /\      /    \
        DET  N    V       NP
         |   |    |      /  \
        the fish like   N     S
                        |    / \
                        it Comp   S
                           / \  /   \
                        for  to NP   VP
                                / \   |
                              DET  N  V
                               |   |  |
                              the fish swim
```

Figure 2-2: English parse tree

```
                S
              / | \
         NP    VP   ADV
         /\     |    |
      DET  N    V   gern
       |   |    |
     die Fische |
            schwimmen
```

Figure 2-3: German parse tree

An English syntactic analyzer, given the source text, would produce a parse tree something like the tree in Figure 2-2[6]. However, a German syntactic generator would require something like the tree in Figure 2-3 in order to produce the translation. As the figures illustrate, the structure of the tree from the English analysis is quite different from the structure of the tree needed to produce the German translation of this sentence. The English verb, "to like," must be transformed into an adverb, "gern" (gladly), and thus the main verb of the German sentence is "schwimmen" (swim). The transformation between the parse tree and the tree necessary for the generation phase, then, is the task which must be performed by the transfer phase.

Structural transfer rules take on a form similar to transformational rules (Chomsky, 1965), in that they transform tree structures to other tree structures. For this example, the transfer rule would take as input a tree with the verb "like" as the verb under an S node, followed by an S node beginning with a Comp consisting of "for" and "to." The transfer rule would output a tree with the German equivalent of the infinitive as the main verb under the S node, with the adverb "gern" modifying the verb. The transfer rule is

[6]This analysis is based on syntactic analyses of similar sentences involving infinitival clauses presented in (Akmajian and Heny, 1975). The exact parse would vary depending on the particular grammar used in the system.

shown in graphic form in Figure 2-4.

```
"LIKE" TRANSFER RULE

Input:                      Output:

     S                             S
   /   \                         / | \
  NP     VP                    NP  VP  ADV
  (1)   /    \                 (1) (2)  |
       V       NP        =>            gern
       |     /   \
     like   N      S
            |    /   \
            it  Comp    S
               /  \   /  \
            for   to NP   VP
                     (1)  (2)
```

Figure 2-4: Transfer rule for translating "to like" into German

2.2.2 Semantic Additions to Syntactic MT

Semantic Features

Many present-day syntax-based machine translation systems make use of a limited amount of semantics during the translation process. Much of this semantic information takes the form of *semantic features* (Chomsky, 1965) (Katz and Fodor, 1963). Examples of MT systems which use semantics features are the METAL system (Slocum and Bennett, 1982), TAUM-METEO (Chandioux, 1976), and ARIANE, the GETA (Grenoble) system (Boitet and Nedobejkine, 1981).

Semantic features are binary features such as +-animate, +-countable, etc., which are attached to lexical items. These features are used to resolve ambiguities which purely syntactic rules cannot by themselves resolve. For instance, consider the following sentences:

John ate the cake with a fork.
John ate the cake with chocolate frosting.

Syntactically, it is ambiguous as to whether the prepositional phrases in these two sentences should be attached to the verb "ate" or the noun phrase "the cake." However, the ambiguity can be resolved with the use of binary semantic features. The word "fork"

is assigned a feature such as +instrument, while the word "frosting" is assigned a feature such as +edible. Then, selectional restriction rules (Katz and Fodor, 1963) are attached to the words "ate" and "cake" which perform checks on these semantic features. The selectional restriction rule attached to "ate" states that the object of the preposition "with" must have the semantic feature +instrument in order for the prepositional phrase to be attached to "ate." Similarly, the selectional restriction rule attached to "cake" states that the object of "with" must have the semantic feature +edible in order for the prepositional phrase to be attached to "cake." Given these semantic features and selectional restriction rules, the structural ambiguity in these two sentences can be resolved.

Binary semantic features as they were used in linguistic theories such as transformational grammar (Chomsky, 1965) were much more limited than they are in current MT systems. In transformational grammar, only a limited number of basic features, such as +-animate, +-countable, etc., were thought to be necessary. However, the use of semantic features in syntax-based MT systems has gone far beyond this original intention. Some systems, such as TAUM-METEO (Chandioux, 1976), have quite large and elaborate sets of semantic features, which map out the semantics of a very limited domain in a fair amount of detail. TAUM-METEO has a detailed set of semantic features for the domain of weather information. Such detailed semantic features were found to be necessary in order to help in the resolution of ambiguities.

Logical Relations

Some syntax-based MT systems also use a limited set of basic logical relations to augment the representations produced by their syntactic analyzers. Examples of such systems are ARIANE (Boitet and Nedobejkine, 1981) and the EUROTRA system (King, 1981).

In these systems, another stage is added to the analysis process, which uses the results of the syntactic parse to assign basic logical relations, such as AGENT, PATIENT, etc., between constituents in the parse tree. This is usually done using a simple mapping between syntactic positions and logical relations. For example, the subject of a verb is usually its AGENT, the object of a verb its PATIENT, etc. Other logical relations are assigned on the basis of prepositions: the INSTRUMENT case would be flagged by the preposition "with," etc.

Although the addition of logical relations to a syntactic representation adds more semantic information, this by no means constitutes a "deep" semantic analysis of the input text. First, logical relations in these systems connect lexical items, not representational structures. Second, the nature of the rules which assign these logical relations is still largely syntactic. By this, I mean that they are restricted to using information about the syntactic structure of the sentence, and semantic features of the lexical items in the sentence. Thus, logical relation assignment rules are very much like syntactic transfer rules.

2.3 Word Disambiguation Using Syntax and Semantic Features

It is crucial for a machine translation system to be able to do word disambiguation. Ambiguous words can often be translated in any of several different ways. This is because there is usually no equivalently ambiguous word in other languages. For one meaning of the ambiguous word, one particular word might be used in another language, but for another meaning of the original word, some other word in the second language might be more appropriate.

Although I will shortly argue that syntactic systems cannot handle the problem of lexical ambiguity in general, it is possible in the syntactic MT paradigm to write lexical transfer rules for some ambiguous words which can choose the correct translation for at least some of the meanings of these words. This is because different meanings of some ambiguous words are used with different syntactic constructions or use words with different semantic features in particular syntactic roles. Thus, rules can be written for these words which examine syntactic constructions or semantic features of various syntactic constituents which choose the correct translation.

An example of such a word is the verb "to leave," which can be translated into Spanish as "salir" or "dejar" (there are other translations of "to leave," but for the purposes of this example we will only consider these two translations). "Salir" means to leave a place, whereas "dejar" means to leave an object at a particular place. Usually, these two meanings can be distinguished by syntactic construction, or by checking semantic features of words in particular syntactic positions, as these sample translations show:

English: They ARE LEAVING for Chicago today.
Spanish: Ellos SALEN para Chicago hoy.

English: I LEFT my book at home.
Spanish: DEJE mi libro en casa.

English: I LEFT my house this morning at six.
Spanish: SALI de mi casa esta manana a las seis.

In the first example, "to leave" is used intransitively, with the preposition "for" following the verb. This syntactic construction is almost never used with the "dejar" meaning of "to leave," so it would be simple to construct a syntactic rule which chose the correct translation for this example. The second and third examples have the same syntactic constructions, but it is still possible to distinguish the two senses of "to leave" by checking the semantic features of the direct object of the verb. The word "house" and other locational nouns could be marked by some semantic feature such as +locational, and then a transfer rule could be written which would use this semantic feature to choose the correct translation of "to leave."

2.4 Problems with Syntax-based MT

2.4.1 Previous Criticisms of Syntax-based MT

Although there are examples of words, such as "to leave," which can be disambiguated using syntactic methods, this is not generally the case. There are many ambiguities that occur in natural language that cannot be resolved using syntactic techniques.

The argument that syntax-based MT systems are not capable of consistently producing high-quality translations due to their inability to effectively deal with ambiguities has been made before. In addition to Bar-Hillel's (1960) argument, discussed earlier in this chapter, more recently (Carbonell, Cullingford, and Gershman, 1978) also argued that FAHQT would not be possible without a deep conceptual analysis of the input text, including the use of scripts and other high-level knowledge structures. One of their examples was the following story:

> John went into a restaurant. He ordered a hamburger. When the hamburger came, he ate it.

If this story were translated into Russian, one would have to use the Russian verb for "to serve" instead of "to come" in the last sentence. Carbonell et. al. demonstrated that syntactic rules could not suffice to choose the Russian verb for "to serve" in this context. Instead, the rules for choosing "to serve" would have to rely on knowledge about the restaurant domain which could only be accessed through a reasonable understanding of what was happening in the story. The rule necessary here would be something like the following: "If something which is normally served by the waiter arrives at the location of the customer, then use the Russian verb for 'to serve' to express that arrival." In order to use this rule, then, a translation system would have to have detailed knowledge about restaurants, such as what items the waiter is likely to bring to the customer, where the customer is likely to be, etc.

2.4.2 A Criticism of Syntax-based MT with Semantics Added

Although some present-day syntax-based MT systems have added limited amounts of semantics, such as semantic features and logical relations, this by no means constitutes the deep conceptual level analysis which Carbonell *et. al.* argued for. I will now argue that, as might be inferred, these limited semantic additions are not enough to overcome the problems involved with translating ambiguous words.

Earlier I showed an example of an ambiguous word, "to leave," for which it was possible to write syntactic transfer rules which chose the correct translation of the word. This was because the syntactic constructions of "to leave" varied, at least sometimes, according to the meaning of the word. If "to leave" was used intransitively, with the preposition "for" following the verb, this syntactic construction almost always corresponded to the "salir" meaning of "to leave," or "to leave a place." Transitive uses of the verb could be distinguished by semantic features of the direct object.

In general, however, word disambiguation is not so easy. With many words, it is difficult to find syntactic properties which distinguish between meanings. An example of such a word is the Spanish verb "ganar" (or its reflexive form, "ganarse"), which can be translated into English as either "to earn" or "to win." Sometimes, syntactic phenomena or semantic features can be used to choose the correct translation, as in the examples below:

Spanish: Yo GANE el aumento de sueldo porque trabaje duro.
English: I EARNED a raise because I worked hard.

Spanish: Yo GANE el juego de poker anoche.
English: I WON the poker game last night.

For these cases, semantic features could be designed which would distinguish the senses of "ganar" on the basis of its direct objects, "aumento de sueldo" (raise) and "juego de poker" (poker game). However, the direct object of "ganar" cannot always provide the information to distinguish between its meanings:

Spanish: En el casino, yo GANE mil dolares en la noche del ano nuevo.

English: At the casino, I WON one thousand dollars on New Year's eve.

Spanish: En el casino, los talladores SE GANARON mil dolares cada uno en la noche del ano nuevo.

English: At the casino, the dealers each EARNED a thousand dollars on New Year's eve.

In these two examples, we see that not only are the direct objects the same, but there appear to be no other straightforward syntactic roles which distinguish them, either. The subjects of the two examples are different, but it is only in conjunction with the prepositional phrase "en el casino" (because we know that dealers work at casinos) that this allows us to distinguish between the two senses of "ganar." A selectional restriction rule to choose the correct translations for these examples would have to check semantic features of the subject of the sentence and the object of the preposition "en."

The examples above can be reworded slightly so that such a rule would not work, either:

Spanish: Los talladores que trabajaron en la noche del ano nuevo en el casino GANARON mil dolares cada uno.

English: The dealers who worked on New Year's eve at the casino each EARNED one thousand dollars.

To handle this example, yet another transfer rule would be needed, to check the subject of the sentence and the prepositional phrase following a verb within a relative clause which follows the subject.

There are even worse examples, such as the following:

Spanish: Despues de trabjar, el tallador en el casino GANO mil dolares en el juego de poker.

English: After working, the dealer at the casino WON one thousand dollars in a poker game.

This time, even though the same cues are in the sentence (a dealer in a casino) which indicate that "ganar" should be translated as "to earn," the additional information that this is occurring after work indicates that the correct translation is "to win."

In short, the number of possible syntactic roles which would have to be checked by transfer rules for the verb "ganar" to ensure the correct translation in all cases is quite large. Almost any syntactic constituent in a sentence could determine which translation of "ganar" would be appropriate. The number of transfer rules needed to check the semantic features of all of these syntactic roles would be enormous.

"Ganar" can only be translated in two possible ways (at least, we have only considered two possible translations of the word). Yet, it is difficult, if not impossible, to write the syntactic transfer rules necessary to distinguish between its two possible

translations. The problem is even worse for very vague or general words or phrases which can be translated in many different ways. For example, in Spanish there is a very vague phrase, "hacer diligencias." Literally, the phrase means "to do diligent actions." In many contexts, it can be translated as "to run errands," as in the following:

Spanish: Maria no puede ir a la reunion porque tiene que HACER MUCHAS DILIGENCIAS.

English: Mary cannot go the gathering because she HAS TO RUN A LOT OF ERRANDS.

In many contexts, however, it is not appropriate to translate "hacer diligencias" as "to run errands." This is because often the context in which this phrase appears allows the reader to infer quite specifically what action it refers to. In these cases, it is often more appropriate in English to use more specific words to describe the action. Because of this, the number of possible translations of "hacer diligencias" is countless, since the number of possible actions to which it could refer is very large. Here are examples in which "hacer diligencias" must be translated differently (sometimes the verb "realizar" (to realize or achieve) is used in place of "hacer"):

Spanish: Juanita salio a HACER UNAS DILIGENCIAS AL MERCADO.

English: Juanita went TO SHOP FOR GROCERIES.

Spanish: Va a pintar su apartamento? -- Si, pero antes tengo que HACER UNAS DILIGENCIAS PARA VER si consigo la pintura que quiero.

English: Are you going to paint your apartment? -- Yes, but first I have TO GO SEE if I can find the paint that I want.

Spanish: La policia REALIZA INTENSAS DILIGENCIAS PARA CAPTURAR a un reo.

English: The police ARE UNDERTAKING AN INTENSE INVESTIGATION in order to capture a criminal.

From these examples, we see that many, many actions can be expressed in Spanish using "hacer diligencias." Because of this, this phrase can be translated into English in many different ways. In the above examples, "hacer diligencias" has been translated as "to run errands," "to shop," "to go," and "to investigate." There are contexts in which it would be translated in many other ways.

There are difficult problems in handling vague words or phrases such as this with transfer rules which rely on syntactic information and semantic features. To see this, consider the transfer rule that would be needed choose the correct translation of "realizar diligencias" in the police story example above, in which "realizar diligencias" is translated as "to investigate." At first glance, one might think that it would be sufficient to check the subject of "realizar diligencias" for a semantic feature such as +authority or +police. However, this is not the case, as the following example illustrates:

Spanish: La reina Isabela va a visitar a la ciudad de Nueva York el lunes. La policia REALIZA DILIGENCIAS para insurar su seguridad durante la visita.

English: Queen Elizabeth will visit New York city on Monday. The police ARE

TAKING PRECAUTIONS to insure her safety during her visit.

In order to determine what translation of "realizar diligencias" is appropriate, other portions of the sentence must also be checked. To see what other parts of the sentence are relevant, consider the line of reasoning that a human reader might follow in order to infer that "realizar diligencias" refers to a police investigation in the earlier example. First, since the prepositional phrase "para capturar" (in order to capture) follows "realizar diligencias," a human reader knows that the action expressed by "realizar diligencias" somehow will lead to a capture, or that the capture is the goal of the "diligencias." Capturing something involves getting control of it, and we know that before we can get control of an object, we have to know where it is and we have to find it. This indicates that perhaps "realizar diligencias" refers to some sort of finding. But when police are trying to find something in order to get control of it, they usually do a formal type of search, or an investigation. Therefore, we know that in this case, "realizar diligencias" refers to a police investigation.

From this line of reasoning, we can see that a great deal of the context surrounding "realizar diligencias" is important in the determination that "to investigate" is the appropriate translation of the phrase. A transfer rule capable of determining the correct translation of "realizar diligencias" in this example would need to check the semantic features of all the items referred to in this line of reasoning. This includes the police ("policia"), the capture ("capturar"), the relation between "diligencias" and "capturar" (given by the preposition "para"), and the criminal (reo). Thus, the transfer rule would need to check the semantic features of all these words. Therefore, the rule would be the following: If "realizar diligencias" appears in a sentence, its subject has the semantic feature +authority, it is followed by a prepositional phrase consisting of "para" followed by an infinitive with the semantic feature +capture, and the direct object of this infinitive has the semantic feature +criminal, then translate "realizar diligencias" as "to investigate."

Unfortunately, in addition to being quite complex, this rule is very example-specific. It depends on the appearance of the appropriate semantic features attached to words which appear as the subject, object of a preposition, and object of an infinitive in the sentence. But the sentence could just as easily have been worded quite differently, and "realizar diligencias" would still be translated in the same way. For example, here is a story which is quite similar in content, but very different in its wording:

Spanish: INTENSAS DILIGENCIAS POR PARTE DE LA POLICIA resultaron en la captura de un reo.

English: AN INTENSE POLICE INVESTIGATION resulted in t arrest of a criminal.

Here, the same line of reasoning as above still applies, but since the various words involved in the line of reasoning ("policia," "captura," and "reo") appear in different syntactic positions, a different transfer rule would be needed to handle this example. This time, the necessary transfer rule would have to be something like the following: If "diligencias" appears as the subject of the verb "resultar," and "diligencias" is followed by a prepositional phrase consisting of "por" followed by a noun phrase with the semantic feature +authority, and "resultar" is followed by a prepositional phrase consisting of "en" followed by a noun phrase with the semantic feature +capture, and this noun phrase is modified by a prepositional phrase consisting of "de" followed by a noun phrase with the

semantic feature +criminal, then translate "diligencias" as "investigation."

This rule is completely different from the transfer rule for the original sentence fact, for most rewordings of the original sentence, the system would have to rely o different transfer rule to correctly translate "diligencias." Thus, since the number syntactic constructions in which "realizar diligencias" could be used to m "investigation" is very large, the number of transfer rules required to translate this ph as "investigation" would also be very large.

So we see that the addition of semantic features does not solve the probl associated with lexical disambiguation. Very vague words which have many diffe senses and which can be translated in many different ways would require a horrendo large number of rules for their disambiguation. The above example of "real diligencias" suggests not only that many rules would be required simply because of large number of ways in which vague words could be translated, but also that e possible translation of a vague word would itself require many rules to cover all possible syntactic variations for which the meaning of the word corresponding to t particular translation could be expressed.

The addition of logical relations in syntax-based MT systems is also inadequate handle the problems of translating ambiguous words. Recall that in some syntax-ba systems, an additional stage is added to the analysis procedure, which is responsible adding simple logical relations to the syntactic parse tree. Just as with the transfer r I have discussed thus far, the rules in this stage can only make use of the synta construction of the text and the semantic features of the lexical items in order to as logical relations.

One might think that these logical relations would provide enough information to down on the number of rules necessary to perform word disambiguation. However, th not the case. Because of the dependence of the rules which assign logical relations syntactic information, the addition of logical relations simply moves the problem from transfer phase to the logical relation assignment phase. To see this, let us assume t the logical relations in Figure 2-1 could be assigned during the analysis of the p investigation example. Given these relations, it appears that the countless rules neces before have been reduced to only one transfer rule, which would be the following: "realiza diligencias" appears in a sentence, its AGENT is a NP with the semantic fea +authority, and it is IN-SERVICE-OF a "capturar" whose PATIENT is an NP with semantic feature +criminal, then translate "realiza diligencias" as "investigate."

This rule would also handle the rewording of the story presented earlier:

Spanish: INTENSAS DILIGENCIAS POR PARTE DE LA POLICIA resultaron en la captura de un reo.
English: AN INTENSE POLICE INVESTIGATION resulted in the arrest of a criminal.

Assuming the analyzer were capable of assigning the correct logical relations, same relations would be assigned as above. So indeed, this addition of logical relat seems to have reduced the number of transfer rules needed to translate "real diligencias."

However, this reduction in the number of transfer rules has been based on assumption that it is possible to design rules which, relying only on syntactic struc and semantic features, could assign the correct logical relations. It this a reason assumption? Consider the police investigation stories again. The original wording of

```
       policia
          |
        AGENT
          |
 realiza diligencias
          |
    IN-SERVICE-OF
          |
       capturar
          |
       PATIENT
          |
         reo
```

Figure 2-1: Logical relations for police investigation story

story was:

> Spanish: La policia REALIZA INTENSAS DILIGENCIAS PARA CAPTURAR a un reo.
>
> English: The police ARE UNDERTAKING AN INTENSE INVESTIGATION in order to capture a criminal.

The rules necessary to assign logical relations would need to relate the syntactic roles of constituents to their logical roles. For this sentence, the rules would be quite straightforward:

> The subject of "realizar diligencias" is also its AGENT.
>
> If a prepositional phrase consisting of "para" followed by an infinitive is attached to "diligencias," "diligencias" fills the logical role IN-SERVICE-OF of the infinitive.
>
> The direct object of the verb "capturar" is also its PATIENT.

These rules would assign the logical relations AGENT, IN-SERVICE-OF, and PATIENT in the appropriate places in the parse tree, thus allowing for the use of the single transfer rule given above to determine the translation of "realizar diligencias."

However, consider the modification of the story presented earlier:

> Spanish: INTENSAS DILIGENCIAS por parte de la policia resultaron en la captura de un reo que dio muerte a una mujer.
>
> English: AN INTENSE POLICE INVESTIGATION resulted in the capture of a criminal who killed a woman.

The assignment of the same logical relations would not be so easy for this sentence. In particular, the assignment of "reo" (criminal) as the PATIENT of "captura" would be quite difficult. The preposition "de" in Spanish is equivalent to either the English "from" or "of." "De" can refer to any of a number of logical relations, including AGENT, PATIENT, or SOURCE (from). Thus, the rule assigning "reo" as the PATIENT of "captura" cannot be so straightforward, since it depends on the disambiguation of the preposition "de." Selectional restrictions from the word "captura" cannot distinguish

between the possible relations to which "de" could refer, because the semantic features of "reo" would fit the selectional restrictions for any of these roles. It is only in the context of the police performing a capture that the logical role PATIENT can be chosen over the other possible logical roles (AGENT and SOURCE).

Because of this ambiguity, then, the rule for assigning the logical relation PATIENT between "captura" and "reo" would have to be the following: If a prepositional phrase consisting of "de" followed by a noun with the semantic feature +criminal is attached to the word "captura," and an action whose AGENT has the semantic feature +authority has been assigned to be IN-SERVICE-OF the word "captura," then the prepositional object of "de" is the logical PATIENT of the word "captura."

This rule is quite complicated, and very specific to this example sentence. It relies on a large number of checks for the appearance of the right words in the right syntactic or semantic roles. Thus, it indicates that the number of rules required to disambiguate the preposition "de" in all contexts would be horrendously large, just as was the case with transfer rules. This is because the appropriate semantic information necessary for the disambiguation of "de" could be found in almost any syntactic role in the sentence. Thus, a hopelessly large number of rules would be necessary to cover all the possible syntactic roles in which the semantic information could appear.

So we see that although at first glance it appeared that the addition of logical relations to the syntactic parse tree greatly reduced the number of transfer rules needed, thus solving the problem, in reality the problem had merely been moved from the transfer phase to the analysis phase. Given the syntactic nature of the logical relation assignment rules, the same problems must arise in constructing these rules as arose before in constructing transfer rules. Since the analysis phase now has the task of assigning logical relations, using only syntactic information and semantic features, the analyzer would require countless rules to assign these logical relations. As a result, the accurate assignment of logical relations within the paradigm of syntax-based analysis is not feasible. Thus, the desired reduction in the number of transfer rules cannot be achieved.

2.5 Conclusion

Syntactically-based machine translation systems are capable of producing output that can be of limited use in practical applications (e.g., TAUM-METEO (Chandioux, 1976), SYSTRAN (Toma, 1977)). However, all such systems either run in a highly restricted domain, in which the problems of lexical ambiguity are highly constrained, or else require a heavy amount of postediting.

Although it might be tempting to hope that syntax-based MT can progress to require smaller amounts of postediting, I have argued that syntax-based MT systems cannot in general solve the problems associated with translating ambiguous words. The large number of syntactic roles in which the information relevant to choosing the correct translation of an ambiguous word can be found results in hopelessly complicated syntactic word disambiguation rules. The use of logical relations in addition to semantic features also fails to solve this problem. Instead, the problem is simply shifted from the transfer phase to the phase responsible for assigning the logical relations. Again, a large number of rules would be necessary to correctly assign the logical relations in cases involving vague or ambiguous words.

The problems encountered in syntax-based approaches to machine translation suggest

that a more semantic approach to the task is needed. In the police investigation examples, it was not possible to write a compact set of transfer rules for "realizar diligencias." This was because the syntactic nature of these transfer rules did not accommodate the encoding of the semantic information which was needed to determine the meaning, and thus the correct translation, of "realizar diligencias." In the examples, a human reader would know that "realizar diligencias" meant "to investigate" because of a piece of world knowledge: a police investigation is an action which the police often perform in service of the goal of arresting someone. Since the stories provided the information that the police were performing some action in service of an arrest, the reader could use this world knowledge to infer that the action, referred to by "realizar diligencias," was a police investigation. A syntax-based machine translation system needs equivalent knowledge in order to determine the correct translation of "realizar diligencias." However, it is very difficult to encode this knowledge in terms of syntactic transfer rules.

The problem which we have encountered with syntactic transfer rules, then, is that some of the knowledge necessary for translation is better represented at a different level. Since the knowledge necessary for the translation of "realizar diligencias" is a piece of world knowledge, it is better represented as such, rather than in terms of syntactic information, since many different syntactic constructions can be used to mean the same thing. This implies that other levels of knowledge should be added to a machine translation system if it is to be able to correctly translate ambiguous words. The system needs more semantics and more world knowledge, and a deeper level of semantic analysis needs to be performed on input texts.

3. Using Semantics for Word Disambiguation

3.1 Introduction

In the last chapter, I attempted to construct transfer rules based on syntactic information and semantic features which would correctly translate ambiguous or vague words. These attempts failed, because of the unmanageably large number of rules that would have been necessary to check all the possible syntactic roles and semantic features which could affect the translation of an ambiguous or vague word. The use of logical relations in addition to semantic features appeared at first to solve this problem. The number of transfer rules that were needed using logical relations was much smaller. However, when I attempted to construct the syntactic rules that would be needed to assign logical relations in the syntactic parse tree, it became obvious that again an unmanageably large number of rules would be needed.

If the problems involved with translating ambiguous or vague words are to be solved, a way must be found to encode the knowledge necessary to translate vague or ambiguous words in a much more compact and manageable way. Using syntax-based techniques, word disambiguation knowledge was encoded in the form of lexical transfer rules or in the form of logical relation assignment rules. Both of these types of rules could only reference syntactic constructions and semantic features. Unfortunately, these encodings resulted in an unmanageably large number of rules. Thus, another encoding of disambiguation knowledge must be found which results in drastically fewer rules.

What sort of encoding of disambiguation knowledge is likely to work? The obvious answer to this question is that a more semantic encoding should have more success. If these rules could be written so that they could rely on "deeper" semantic information, rather than on syntactic information and semantic features, then perhaps the number of rules could be reduced.

Consider the police investigation stories from the last chapter:

Example 1

Spanish: La policia REALIZA INTENSAS DILIGENCIAS para capturar a un reo.

English: The police ARE UNDERTAKING AN INTENSE INVESTIGATION in order to capture a criminal.

Example 2

Spanish: INTENSAS DILIGENCIAS por parte de la policia resultaron en la captura de un reo.

English: AN INTENSE POLICE INVESTIGATION resulted in the capture of a criminal.

Two completely different syntax-based transfer rules were needed to facilitate the correct translation of "diligencias" for these two sentences. But when logical relations were

added to the syntactic parse tree, this allowed for the same transfer rule to be used for both of these sentences. Intuitively, this seemed to make sense; although the two syntactic rules did not look very similar on the surface, they both reflected the combination of the words "diligencias," "capturar," "policia," and "reo" in semantically the same way. Both rules encoded the fact that whenever the police are the actors of an action which is in service of or leads to the capture of a criminal, that action is probably an investigation.

The problem with the use of logical relations as they were used in syntactic MT systems was that the rules which assigned logical relations still only had access to syntactic information and semantic features. As a result, when there was an ambiguity to resolve during the assignment of logical relations, the number of rules needed to resolve the ambiguity was again unmanageably large. Given the fact that writing transfer rules was easier once more semantic information was available, perhaps the problems involved in assigning the logical relations could also be overcome if more semantic information were available during logical relation assignment, also.

In this chapter, I will argue that the encoding of word disambiguation knowledge in terms of "deeper" semantic information does indeed result in the need for drastically fewer rules in order to perform the translation of ambiguous or vague words. This includes the logical relation assignment rules as well as the transfer rules themselves. I will define what I mean by "deeper," in terms of the kind of semantic information necessary to accomplish this rule reduction. I will argue that such a reduction requires the use of high-level knowledge structures which must be independent of lexical items, and that a semantic representation, also distinct from the lexical items in a source text, must be built during the parsing process in order for these "deeper" semantic knowledge structures and disambiguation rules to be used.

3.2 The Need for Semantic Representations

It is possible to reformulate the logical relation assignment rules which I discussed in the last chapter in a more semantic way. Instead of relying on syntactic information and semantic features, these rules can be rewritten in in terms of "deeper" semantic information in a more efficient way, thus resulting in a drastic reduction in the number of rules needed. To see this, let us examine the police investigation stories above and see how the logical relation assignment rules from before can be rewritten.

Recall that the following rules were capable of assigning the correct logical relations for example 1 above:

> Rule 1A: The subject of "realizar diligencias" is also its logical AGENT.
>
> Rule 1B: If a prepositional phrase consisting of "para" followed by an infinitive is attached to "diligencias," "diligencias" fills the logical role IN-SERVICE-OF of the infinitive.
>
> Rule 1C: The object of the verb "capturar" is also its logical PATIENT.

Logical relation rules for example 2 are not as straightforward. In this sentence, the AGENT of the "diligencias" appears in the prepositional phrase beginning with "por" (by) following "diligencias," the logical relationship IN-SERVICE-OF between "diligencias" and "captura" is expressed by the verb "resultaron," and the PATIENT of

"captura" appears in the prepositional phrase beginning with "de" (of or from) which follows "captura."

The syntax-based rules to assign these logical relations for this sentence would have to be something like the following:

Rule 2A: If a prepositional phrase consisting of "por" followed by a noun is attached to "diligencias," the noun is the AGENT of "diligencias."

Rule 2B: If a prepositional phrase consisting of "en" followed by a noun is attached to the verb "resultar," then the subject of "resultar" fills the logical role IN-SERVICE-OF of the object of the preposition "en."

Rule 2C: If a prepositional phrase consisting of "de" followed by a noun with the semantic feature +criminal is attached to the word "captura," and an action whose AGENT has the semantic feature +authority has been assigned to be IN-SERVICE-OF the word "captura," then the prepositional object of "de" is the logical PATIENT of the word "captura."

The last of these three rules, rule 2C, is quite a bit more complicated than the corresponding rule which assigned "reo" as the AGENT of the "captura" for example 1. Recall from the last chapter that this was due to the fact that the Spanish preposition "de" (of or from) can refer to many possible logical roles, such as AGENT, PATIENT and SOURCE. Selectional restriction rules from the word "captura" cannot distinguish between these roles, since "reo" (criminal) could conceivably be the logical AGENT, PATIENT, or SOURCE of the word "captura." It is only in the context of the police performing a capture that the logical role PATIENT can be chosen over the other possible roles (AGENT and SOURCE), since police tend to capture criminals.

How can these rules be rewritten in such as way as to simplify the third rule above? To answer this question, consider the following example:

Example 3

Spanish: Intensas diligencias por la policia resultaron en la ARRESTA de un reo.

English: An intense police investigation resulted in the ARREST of a criminal.

For this sentence, the rule which would be needed to assign "reo" (criminal) as the logical PATIENT of "arresta" would be simpler, and much more general. This is because the selectional restriction information from the word "arresta" is more specific than the selectional restriction information from the word "captura." Selectional restriction rules from "arresta" would restrict the PATIENT of "arresta" to have the semantic feature +criminal, since the people who police arrest are criminals. This is a more specific selectional restriction than was provided by the word "captura," since criminals are not the only things that are captured. The more specific selectional restriction of "arresta" can be used to help disambiguate "de" in this context, by matching the semantic features of the object of the preposition "de" with the selectional restrictions of the various logical roles which "de" can refer to. Thus, for example 3 above, the logical relation assignment rule for "de" would be the following:

Rule 3C: The preposition "de" can refer to one of three roles: AGENT, PATIENT, and SOURCE. If a prepositional phrase consisting of "de"

followed by its prepositional object, word A, is attached to word B, and if the semantic features of word A match the selectional restrictions of word B for one of the logical roles which "de" can refer to, then assign word A to fill that logical role of word B.

This rule would assign "reo" to be the PATIENT of "arresta" as follows: "Reo" is the object of the preposition "de" in example 3. Since this prepositional phrase is attached to "arresta," and since the semantic feature +authority of the word "reo" matches the selectional restriction for the PATIENT of "arresta," which is one of the roles which "de" can refer to, "reo" would be assigned to be the PATIENT of "arresta."

Comparing this rule with the rule which was needed for example 2, we see that this rule is simpler by virtue of the fact that it does not need to check for an action whose AGENT has the semantic feature +authority assigned as IN-SERVICE-OF the "arresta." It is also much more general than the rule which was needed for example 2, since it is not specific to the context of "de" appearing after the word "arresta." The same rule could be used in any context in which "de" follows a word whose selectional restrictions match the semantic features of the prepositional object of "de."

A simpler and more general rule can be used for example 3 because of the more specific selectional restrictions of "arresta." These selectional restrictions are specific enough to determine that "reo" (criminal) must be the PATIENT of the "arresta," since the selectional restrictions for the PATIENT of "arresta" match the semantic features of the word "reo." However, this match could not be made for the word "captura," since its selectional restrictions were not as specific.

Is it possible to make it so that the more general rule used for "de" following "arresta" could also be used to disambiguate "de" appearing after "captura"? We can answer this question by considering how a person might understand example 2 above. In the context in which "de un reo" appears in example 2, it is clear to a human reader that "reo" must be the PATIENT of "captura," and therefore "de" means PATIENT. This is because a human reader can infer that it is likely that the capture in this sentence is being performed by the police, since the police are performing an action which is in service of the capture. Since it is the police who are performing the capture, it is likely that the capture is actually an arrest, since an arrest is the type of capturing that the police are most likely to do. In other words, a human reader can easily infer that "captura" in this context means "arrest." Because of this, the more specific selectional restrictions of words that mean "arrest" can apply in this context, thus providing enough information to infer that the criminal is the object of the arrest.

This line of inferencing suggests that the simpler, more general rule used to disambiguate "de" in example 3 can indeed be used to disambiguate "de" in example 2, if a new type of rule is introduced into the system. What is needed is a rule which can *change* the selectional restrictions of the word "captura" in this context, reflecting the fact that "captura" in this context most likely refers to an arrest. This rule would be something like the following:

Rule 3D: If a verb whose AGENT has the semantic feature +authority fills the logical role IN-SERVICE-OF of the word "captura," then use the selectional restrictions that usually are used with the word "arresta" (arrest) for the word "captura."

This rule together with rule 3C above would be able to disambiguate "de" in example

2. This is because the more specific selectional restrictions of the word "arresta" would provide the additional information needed to use this rule. Rule 3C would assign "reo" to be the logical PATIENT of "captura," because the context in example 2 would satisfy the conditions of Rule 3D, and thus the selectional restrictions normally associated with the word "arresta" would be in effect for the word "captura."

At first glance, the reformulation of rule 2C, the original disambiguation rule for "de" in example 2, in terms of rules 3C and 3D does not look any better than the original rule. Instead of having one rather complex rule for for determining the meaning of "de" in example 2, now two rules are required. So at first glance, all I have done is to replace one rather complex rule with two equally complex rules.

However, the important difference between rule 2C and rules 3C and 3D is that the original single rule was a *special-purpose disambiguation rule*, while the reformulation is written in terms of two more general rules. By this I mean that the only purpose of rule 2C was to perform the disambiguation of the word "de" in the context in which the word appeared in example 2 (or other very similar contexts). This rule could not be used for any other purpose. In contrast, rules 3C and 3D are more general, in that they could be used in other contexts, also. For example, rule 3C can also be used to disambiguate "de" in example 3. The same rule could be used in many other contexts, since it does not depend on the appearance of the word "captura" before the word "de," as did rule 2C. Rule 3D is also more general than was the original rule, in that it useful in other contexts. For instance, rule 3D could be used in conjunction with rule 3C to disambiguate "de" in the following example:

Spanish: Intensas diligencias por la policia resultaron en la primera captura DEL policia mas nuevo de la ciudad.

English: An intense police investigation resulted in the first arrest BY the city's newest policeman.

Rule 3D would apply to this sentence, since again an action whose AGENT has the semantic feature +authority is IN-SERVICE-OF a "captura." Then, rule 3C would perform the disambiguation of "de," since the more specific selectional restrictions of the word "arresta" would apply, providing the information that "de" must refer to the logical role PATIENT.

So we see that although this reformulation has resulted in the replacement of one complex rule with two rules, the result is actually a reduction in rules, since the two rules in the reformulation can apply to a wider range of contexts. Thus, the number of rules needed to disambiguate "de" in general has been reduced.

One way to simplify logical relation assignment rules, then, is to allow selectional restriction rules to be dynamic. Instead of one particular static set of selectional restriction rules attached to a word, these selectional restriction rules should be changeable, depending on the context in which a word appears. Another way to say this is that selectional restriction rules *should not be lexically based*. Rather than being tied directly to words themselves, selectional restriction rules should be tied to *meanings of words*. In some contexts, "captura" means the same thing as "arresta." Namely, they both refer to some (non-lexical) concept, ARREST. In these contexts, "captura" and "arresta" should have the same selectional restriction rules, the same selectional restriction rules that any word meaning ARREST should have.

Given that selectional restriction rules should be dynamic, varying according to the meaning of a word in a particular context, we must have some way to keep track of what

selectional restriction rules apply to a word. This means that during the parsing process, the parser should keep track of the meaning of the text that it is parsing, and then use the selectional restrictions from those meanings in order to aid the parsing process. In other words, the parser should keep a *semantic representation* of the text it is processing. This representation is independent of the lexical items in the text, and reflects the text's "deep" semantic meaning. In turn, the representation provides the means for keeping track of what selectional restriction rules are active for a given word.

To make this more concrete, let us continue discussing example 2 above. We can rewrite the rules which changed the selectional restrictions of the word "captura" in terms of rules which operate on a semantic representation, as follows:

CAPTURE selectional restriction rules: The ACTOR of a CAPTURE is a PERSON. The OBJECT of a CAPTURE is a PHYSICAL OBJECT.

ARREST selectional restriction rules: The ACTOR of an ARREST is an AUTHORITY. The OBJECT of an ARREST is a CRIMINAL.

Definition of "captura": Given no contextual information, the word "captura" should be represented by the concept CAPTURE.

Definition of "de": The preposition "de" can refer to the semantic roles ACTOR, OBJECT, and SOURCE. Use selectional restriction rules from the meaning of the word to which "de" is attached and the semantic features of the object of "de" to determine which semantic relation is appropriate in the context.

Contextual CAPTURE inference rule: If an action whose ACTOR is the POLICE fills the semantic role IN-SERVICE-OF of a CAPTURE, then infer that the POLICE are also the ACTORs of the CAPTURE, and that therefore the CAPTURE is actually an ARREST. Use the selectional restriction rules of the concept ARREST to guide attachments to the verb or noun previously represented by CAPTURE.

The definition of "de" above, along with the Contextual CAPTURE Inference Rule, are equivalent to rules 3C and 3D from before. In addition to this, the above set of rules includes the selectional restriction rules which would be attached to the concepts CAPTURE and ARREST. These selectional restriction rules are the same as were attached to the words "captura" and "arresta" before.

This set of rules would choose the correct meaning of the preposition "de" in example 2 as follows: first, the representation of "captura" would be built as CAPTURE, as specified by the definition of "captura." Next, the Contextual CAPTURE Rule would apply, since an action whose ACTOR would be the POLICE would be IN-SERVICE-OF the CAPTURE. This rule would change the representation of "captura" to ARREST, thus making available the selectional restrictions from ARREST. Finally, the information from the definition of "de" in combination with the selectional restrictions from ARREST would provide enough information to determine that "de" refers to the semantic role OBJECT.

Thus, we see that we have reformulated the syntax-based rules for disambiguating the word "de" in terms of rules which require that a semantic representation of the sentence be built during parsing. Given this semantic representation, the rules necessary for the disambiguation of "de" in examples 1-3 above can be written in a more general way, so that they apply to a wider range of contexts than did the original syntax-based

rules. Because of this generality gained in the reformulation of the disambiguation rules, the total number of rules that would be needed to disambiguate "de" in general has been reduced substantially.

3.3 Adding Abstraction Knowledge

So far, the syntax-based logical relation assignment rules have been rewritten in a more general way, using semantic representations. The increase in generality is due to the addition of rules, such as the Contextual CAPTURE Inference Rule, which manipulate the semantic representation built during parsing.

The Contextual CAPTURE inference rule can be improved in such a way that it is even less special-purpose, if we add still more semantic information to the system. Consider the role which this rule is playing. We know that policemen are likely to perform arrests. Therefore, when an action, such as CAPTURE, appears in a story, and the ACTOR of the action is the POLICE, we can infer that it is likely that the CAPTURE is actually an ARREST. This is because CAPTURE is an *abstraction*, or a more general version, of ARREST.

The reason this inference can be performed, then, is because of the *abstraction knowledge* which links the two concepts CAPTURE and ARREST. Namely, we know that CAPTURE is a more abstract version of ARREST, or ARREST is a specific type of CAPTURE. This suggests, then, that disambiguation rules can be written more generally if abstraction knowledge is added to the semantic information in the system. Generalizing from the example involving CAPTURE and ARREST, abstraction knowledge can be useful in parsing because it allows us to make the inference that when an action is known to have a prototypical actor, and that prototypical actor has been assigned to be the ACTOR of an abstraction or more general version of that action, then we can infer that it is likely that the action is actually the more specific action.

Using this generalization, we can reformulate the Contextual CAPTURE Rule to be more general, if we add the appropriate abstraction information to our set of rules:

CAPTURE-ARREST abstraction rule: CAPTURE is an abstraction of ARREST.

Abstraction Inference Rule: If concept B is the ACTOR of action A, and B is the prototypical ACTOR of action C, and action A is an abstraction of action C, then infer that in this case action A is really action C. Use the selectional restriction rules of the concept C to guide attachments to the verb or noun previously represented by concept A.

Now, the Contextual CAPTURE Inference Rule has been rewritten in a much more general way, as the Abstraction Inference Rule. Using this rule, along with the rules from the last section for disambiguating "de," we now have a set of five rules, none of which are specific to the context of this example:

CAPTURE selectional restriction rules: The ACTOR of a CAPTURE is a PERSON. The OBJECT of a CAPTURE is a PHYSICAL OBJECT.

ARREST selectional restriction rules: The ACTOR of an ARREST is an AUTHORITY. The OBJECT of an ARREST is a CRIMINAL.

Definition of "captura": Given no contextual information, the word "captura" should be represented by the concept CAPTURE.

Definition of "de": The preposition "de" can refer to the semantic relations ACTOR, OBJECT, and SOURCE. Use selectional restriction rules from the meaning of the word to which "de" is attached and the semantic features of the object of "de" to determine which semantic relation is appropriate in the context.

Abstraction Inference Rule: If concept B is the ACTOR of action A, and B is the prototypical ACTOR of action C, and action A is an abstraction of action C, then infer that in this case action A is really action C. Use the selectional restriction rules of the concept C to guide attachments to the verb or noun previously represented by concept A.

Whereas before the Contextual CAPTURE Rule only applied to contexts in which the ACTOR of a CAPTURE is the POLICE, the Abstraction Inference Rule is applicable to any situation in which the ACTOR of an action indicates that the action is more specific. Thus, we have reformulated the rules for disambiguating "de," using only general disambiguation rules and rules about the semantics of the concepts which appear in the context.

3.4 Semantic Information and Transfer Rules

So far, I have introduced semantic representations and knowledge about abstraction into the word disambiguation rules. This additional information has greatly reduced the complexity of the rules necessary for the disambiguation of the word "de."

Recall that the problem of writing rules to disambiguate the word "de" came from the desire to simplify the transfer rules for the word "diligencias." If logical relations were used in the transfer rules, the number of rules necessary for the word "diligencias" was reduced. In fact, the same transfer rule could be used for both examples 1 and 2 above. This rule was the following:

> If "diligencias" appears in a sentence, its ACTOR is a NP with the semantic feature +authority, and it is IN-SERVICE-OF a "capturar" whose OBJECT is an NP with the semantic feature +criminal, then translate "diligencias" as "investigate."

However, writing this transfer rule in terms of logical relationships between words in the sentence required that rules be written which could assign these relationships. Attempts to do this using syntactic information and semantic features failed, due again to the large number of rules which were needed for this task. Words like "de," which could refer to a number of logical relationships, required countless rules in order to be correctly disambiguated.

I have succeeded in reducing the number of disambiguation rules for the word "de" by adding semantic representations and abstraction knowledge to the body of knowledge used by the system. Now, the transfer rule above must be combined with this reworked version of the disambiguation rules, so that we have a complete set of rules capable of translating "diligencias" for the examples which we have discussed.

In order to use the new semantic relation assignment rules with the transfer rule for

"diligencias," the transfer rule must also be rewritten. As it stands, the transfer rule refers to semantic features and lexical items. But these semantic features have been replaced in the semantic relation assignment rules by representational items. Thus, the transfer rule must be rewritten in terms of these representational items, also.

To do this, we can introduce another semantic concept, INVESTIGATE. This is the concept which will ultimately represent "diligencias" in these examples. Then, the transfer rule would simply be:

> Transfer rule: A word represented by the concept INVESTIGATE should be translated into English using some form of the word "investigate."

Now the task is to write rules which will cause the representation of the word "diligencias" in the examples to become INVESTIGATE. If we introduce abstraction knowledge once again, this can be done as follows:

> Definition of "diligencias": The word "diligencias" refers to the general concept ACTION.
>
> Definition of "captura": The word "captura" refers to the concept CAPTURE.
>
> FIND selectional restrictions: The ACTOR of FIND is a PERSON. A FIND is often done IN-SERVICE-OF a CONTROL.
>
> INVESTIGATE selectional restrictions: The ACTOR of INVESTIGATE is the POLICE. An INVESTIGATE is often done IN-SERVICE-OF an ARREST.
>
> INVESTIGATE-FIND Abstraction Rule: FIND is an abstraction of INVESTIGATE.
>
> CAPTURE-CONTROL Abstraction Rule: CONTROL is an abstraction of CAPTURE.
>
> FIND-CONTROL Inference Rule: If an ACTION is being done IN-SERVICE-OF a CONTROL, then infer that the action is a FIND.

Here, the transfer rule has been rewritten using the representational and abstraction information introduced in the last two sections. It has been replaced by several rules. First, there is a definition of "diligencias," resulting in a general concept, ACTION, to be used to represent "diligencias" at first. Then, other concepts, FIND and INVESTIGATE are defined. These concepts will also be used to represent "diligencias" along the way. Finally, abstraction knowledge linking FIND and INVESTIGATE, as well as an inference rule telling when to infer that an ACTION is a FIND, has been added.

INVESTIGATE would become the representation of "diligencias" using these rules in the following way: first, ACTION would be the representation of "diligencias." The ACTOR of the ACTION would be assigned to be the POLICE, and this action would be assigned as IN-SERVICE-OF a CAPTURE. Since CAPTURE is a type of CONTROL, the FIND-CONTROL Inference Rule would change the representation of "diligencias" to be FIND. Then, the Abstraction Inference Rule from before would conclude that the representation of "diligencias" should actually be INVESTIGATE, since FIND is an abstraction of INVESTIGATE and POLICE is the prototypical ACTOR of INVESTIGATE.

Again, I have replaced one rule with several, resulting in an apparent increase in the number of rules, not a reduction. However, these rules are all much more general than

the single transfer rule for "diligencias" was. Only the definition of "diligencias" has specifically to do with that word. All the other rules would be applicable in a much wider range of contexts, and thus would be useful in disambiguating other words in semantically similar contexts.

3.5 Scriptal Knowledge and Word Disambiguation

One more change can be made to the rules in the last section, to make them even more general. The FIND-CONTROL rule above reflects the fact that since the action CONTROL is often preceded by the action FIND, and that FIND is often done IN-SERVICE-OF CONTROL, any abstraction of FIND (including a very general abstraction, ACTION) which precedes a CONTROL and is IN-SERVICE-OF the CONTROL may be inferred to be a FIND. In general, then, if one action often precedes a second action, and an abstraction of the first action has already been assigned to precede the second action, a reasonable inference is that the abstraction of the first action actually is the first action.

Thus, the FIND-CONTROL rule can be rewritten in a more general way, as follows:

CONTROL Event Sequence Rule: The action FIND often precedes the action CONTROL. The FIND is done IN-SERVICE-OF the CONTROL.

SCRIPTAL INFERENCE RULE: If action A is part of a known sequence of actions, and action A is mentioned in a story, then expect the mention of other actions in the known sequence, also.

EXPECTED ACTION INFERENCE RULE: If action A is expected in a story, and action B is mentioned in the story, and action B is an abstraction of action A, then infer that action B is actually action A.

These three rules replace the FIND-CONTROL rule in the following way: the CONTROL Event Sequence Rule provides the information that FIND followed by CONTROL is a known sequence of events that commonly occurs. When CAPTURE is built to represent "captura" in the police investigation stories, the Scriptal Inference Rule would cause the action FIND to be expected, since CAPTURE is a type of CONTROL, which appears in the FIND-CONTROL event sequence. Finally, since "diligencias," represented by ACTION, preceded the CAPTURE, the Expected Action Inference rule would cause the representation of "diligencias" to be changed to FIND, since FIND would be an expected action.

In the above set of rules, we find another type of semantic knowledge: *scriptal knowledge*. Scriptal knowledge is knowledge about commonly-found sequences of events. In this case, the sequence consists of two events: FIND followed by CONTROL. This reflects the knowledge that we have that often, in order to get control of something, it has to first be found. Inclusion of this scriptal knowledge allows us to write the FIND-CONTROL rule from before in terms of two extremely general rules: the Scriptal Inference Rule and the Expected Action Inference Rule. These rules are applicable to many contexts, whenever an event which is part of a known sequence of events appears in a story. Thus, these rules are very general, and can aid in the disambiguation of words in a large number of contexts.

3.6 Putting it All Together

We have seen that the addition of several types of semantic information to a parsi system results in the ability to write disambiguation rules in a much more general fashio thus resulting in a vast reduction in the number of disambiguation rules necessary over wide range of contexts. I have advocated the addition of three different types of semant information: semantic representations built during parsing, abstraction information, ar scriptal knowledge. These different types of knowledge have all contributed to tl generality of the disambiguation rules.

Putting together all the rules which I have discussed in the last sections, the followi set of rules will perform the disambiguation of "diligencias" in the examples which I ha discussed:

RULES RELATING SYNTACTIC POSITION TO SEMANTIC ROLE

Rule 1A: The subject of a verb is also often its semantic ACTOR.

Rule 1B: The object of a verb is also often its semantic OBJECT.

Rule 1C: If a prepositional phrase consisting of "para" followed by an infinitive is attached to an ACTION, the ACTION fills the semantic role IN-SERVICE-OF of the infinitive.

Rule 1D: If a prepositional phrase consisting of "por" followed by a noun is attached to an ACTION, the noun is often the ACTOR of the ACTION.

Rule 1E: If a prepositional phrase consisting of "en" followed by a noun is attached to the verb "resultar," then the subject of "resultar" fills the semantic role IN-SERVICE-OF of the object of the preposition "en."

Rule 1F: If a prepositional phrase consisting of "de" followed by a noun is attached to an ACTION, then the object of the preposition is either the semantic ACTOR, OBJECT, or SOURCE of the action.

DEFINITIONS OF WORDS:

Rule 2A: "Diligencias" and "realizar diligencias" are represented by the concept ACTION.

Rule 2B: "Capturar" is represented by the concept CAPTURE.

Rule 2C: "Policia" is represented by the concept AUTHORITY.

Rule 2D: "Reo" is represented by the concept CRIMINAL.

CASE FRAMES OF CONCEPTS:

Rule 3A: The ACTOR of ACTION is a PERSON.

Rule 3B: The ACTOR of FIND is a PERSON. FIND is often done IN-SERVICE-OF a CONTROL.

Rule 3C: The ACTOR of INVESTIGATE is an AUTHORITY.

INVESTIGATE is often done IN-SERVICE-OF an ARREST.

Rule 3D: The ACTOR of CONTROL is a person. The OBJECT of CONTROL is a thing.

Rule 3E: The ACTOR of CAPTURE is a person. The OBJECT of CAPTURE is a thing.

Rule 3F: The ACTOR of ARREST is an AUTHORITY. The OBJECT of ARREST is a CRIMINAL.

ABSTRACTION RULES:

Rule 4A: ACTION is an abstraction of FIND.

Rule 4B: FIND is an abstraction of INVESTIGATE.

Rule 4C: CONTROL is an abstraction of CAPTURE.

Rule 4D: CAPTURE is an abstraction of ARREST.

SCRIPTAL RULES:

Rule 5A: GET is a sequence of actions consisting of FIND followed by CONTROL.

Rule 5B: POLICE-CAPTURE is a sequence of actions consisting of INVESTIGATE followed by ARREST.

INFERENCE RULES:

Rule 6A (ABSTRACTION INFERENCE RULE): If A is the ACTOR of action B, and A is the prototypical ACTOR of action C, and action B is an abstraction of action C, then infer that in this case action B is really action C. Use the selectional restriction rules of the concept C to guide attachments to the verb or noun previously represented by concept B.

Rule 6B (SCRIPTAL INFERENCE RULE): If action A is part of a common sequence of actions (a script), and action A is mentioned in a story, then expect the mention of other actions in the script, also.

Rule 6C (EXPECTED ACTION INFERENCE RULE): If action A is expected in a story, and action B is mentioned in the story, and action B is an abstraction of action A, then infer that action B is actually action A.

Given all these rules, the disambiguation of "realizar diligencias" for the original wording of the police investigation story, or example 1 above, would proceed as follows: at first, the representation of "realizar diligencias" would be ACTION, because of rule 2A. Then, the ACTOR of "realizar diligencias" would be filled in with AUTHORITY, since rule 1A would fill in the subject of a verb as its ACTOR. Next, the representation of

"capturar" would be built as CAPTURE, and the Expected Action Inference Rule (rule 6C) would cause the script GET to be activated, thus also activating expectations that the event FIND should occur in the story (because of rule 5A). The preposition phrase "para capturar" would be attached to "realizar diligencias," and rule 1C would cause the ACTION representing "realizar diligencias" to be placed in the IN-SERVICE-OF role of the CAPTURE. After this, since CONTROL is an abstraction of CAPTURE, rule 6B, the Scriptal Inference Rule, would cause the inference that the representation of "realizar diligencias," ACTION, should be changed to FIND, since ACTION is an abstraction of FIND, and FIND appears in the script GET. Finally, since the representation of "diligencias" is now FIND, and the ACTOR of FIND is an AUTHORITY, the Abstraction Inference Rule (rule 6A) would cause FIND to be changed to INVESTIGATE, since FIND is an abstraction of INVESTIGATE. Thus, the phrase "realizar diligencias" would be translated as "to investigate."

For the second wording of the sentence, example 2 above, the disambiguation of "diligencias" would proceed as follows: again, at first the representation of "diligencias" would be the general concept ACTION, because of rule 2A. Next, AUTHORITY would be built as the representation of "policia" because of rule 2C. AUTHORITY would then be assigned to be the ACTOR of the ACTION, due to rule 1D. After the representation CAPTURE of the word "captura" was built, and the script GET was activated by the Expected Action Inference Rule (rule 6C), rule 1E would fill in the ACTION to be IN-SERVICE-OF the CAPTURE. Again, at this point, the Scriptal Inference Rule (6B) would change the ACTION to be a FIND. Next, FIND would be changed to be INVESTIGATE, by the Abstraction Inference Rule (6A), due to the fact that FIND is an abstraction of INVESTIGATE (rule 4B), and AUTHORITY fits the prototype for the ACTOR of INVESTIGATE (rule 3C). This would cause the representation CAPTURE to be changed to ARREST, due to the Scriptal Inference Rule again (6B). Finally, after the building of the representation CRIMINAL for the word "reo," rule 1F in combination with the selectional restriction rules for ARREST (3F) would result in the assignment of CRIMINAL as the OBJECT of the ARREST.

3.7 Conclusion

To review, what have I done here? Originally, two syntax-based transfer rules were needed to perform the translation of "diligencias" in these two examples. It was clear that many more rules would be needed to handle other semantically similar but syntactically different stories. When logical relations were introduced to try to remedy the situation, only one transfer rule was needed, but it was clear that many rules would be needed to perform the assignment of these logical relations in cases where the assignment was ambiguous, such as with the preposition "de." Finally, when I attempted to remedy this situation by introducing more semantics into the logical relation assignment rules, it became clear that the paradigm of syntax-based parsing would have to be changed in order for this to be possible. First, semantic representations were necessary in order to handle non-static selectional restriction rules. Static selectional restriction rules could not provide specific enough semantic information in certain contexts, such as with the word "captura" in a police context. In order to make selectional restriction rules specific enough and sensitive to context, they had to be indexed according to the meanings of words, not the words themselves. Thus, they should be attached to concepts, and representations of the text, consisting of these concepts,

4. A Critique of Previous Work in Conceptual Analysis

4.1 Introduction

In the last chapter, I gave motivation for the use of semantic analysis techniques in machine translation. Thus, I will now examine some of past research in conceptual analysis, to see what relevance it has to the problems which were encountered in syntax-based translation techniques, and to the other machine translation problems which I discussed in chapter 1. In particular, I will disuss request-based parsers, such as Riesbeck's analyzer (Riesbeck, 1975), and the many other similar integrated parsers which have followed.

In conceptual analysis, the goal of the parser is to produce a conceptual (i.e., language-independent) representation of the input text, rather than to produce a syntactic parse tree. Some important issues arise due to this difference in goals. First, the parser must contain a conceptual knowledge base, as well as linguistic or syntactic knowledge, since a conceptual representation must be produced. How should these two types of knowledge be integrated, if at all? Should syntactic knowledge and conceptual knowledge be completely separate, or completely integrated, or somewhere in between? And in processing, should syntactic and semantic analysis proceed in tandem, or should the process be modularized?

Another issue also arises from the different goal of a conceptual analyzer. Since the final product of the parser is no longer syntactic in nature, what should the role of syntax be in conceptual analysis? Is it necessary to build a syntactic representation of an input text during parsing, if this representation is simply going to be thrown away afterward, leaving the conceptual representation? Or can a conceptual analyzer get away with less syntactic analysis, since this analysis is not the final goal of the parser?

In this chapter, I will explore the approaches taken in previous conceptual analysis research with regards to the issues of integration of syntax with conceptual knowledge, and the role of syntactic knowledge. I will also discuss the ways in which these issues relate to machine translation, and the problems which we encountered in the last two chapters with syntax-based transfer rules. I will conclude that problems exist in past conceptual analysis research which are analogous to the problems with syntax-based transfer rules. Just as with the transfer rules, the way in which knowledge is represented in past conceptual analyzers sometimes results in the need for a large number of rules to encode the knowledge necessary for the resolution of syntactic and semantic ambiguities. This problem is analogous to the rule explosion problem with syntax-based transfer rules, because it is due to the knowledge being stored at the wrong level of generality. With transfer rules, disambiguation knowledge had to be encoded in terms of syntactic structure and semantic features. This was an inappropriate level at which to encode this knowledge, since it was often semantic or conceptual in nature. As we will see, an analogous problem occurs with the encoding of syntactic and conceptual knowledge in past conceptual analyzers. Thus, although the problems which were encountered in the previous chapter pointed to the need for more semantics, the existing conceptual analysis research must also be improved upon in order to solve the problems associated with the

should be built during the parse.

These changes resulted in the need for more rules in order to handle these two examples, but the rules were less example-specific, and thus capable of handling many other semantically similar examples. As a result, the horrendously large number of rules needed to disambiguate "diligencias" in all contexts had been reduced. But it was possible to improve these rules further so as to make them still more general. In order to do this, I first introduced *abstraction knowledge*, or knowledge about what concepts were more general versions of other concepts. Finally, I introduced *scriptal knowledge*, or knowledge about common sequences of events. After introducing this knowledge, the only rules needed to perform the disambiguation of "diligencias" were three extremely general rules, which weren't even specific to the problem of disambiguation: the Abstraction Inference Rule, the Scriptal Inference Rule, and the Expected Action Inference Rule. Many other rules were included in the final set of rules which performed the disambiguation of "diligencias," but none of these rules dealt specifically with the problem of disambiguation, either. They consisted of word definitions, concept definitions, rules relating syntactic position to semantic meaning, and definitions of particular scripts.

So by using semantic representations, abstraction knowledge, and scriptal knowledge, and by writing rules in terms of these different kinds of knowledge in addition to rules about syntax, the disambiguation of "diligencias" can be done in these two examples *using NO example-specific rules*. This set of rules is capable of disambiguating "diligencias" to mean "to investigate" in many, many different contexts. Not only that, but many of the rules could be used in the disambiguation of "diligencias" in contexts in which "diligencias" takes on a different meaning, or in the disambiguation of other words in similar semantic contexts. So although the number of rules needed to do these two examples has increased substantially using more semantics, the prospect of rule explosion that seemed inevitable using syntax-based methods has been greatly lessened.

translation of texts which contain potential ambiguities.

4.2 Request-based Parsing

One approach that has been taken to conceptual analysis has involved the use of *requests*, or test-action pairs, to encode parsing knowledge. Requests were first used in Riesbeck's analyzer (Riesbeck, 1975). The requests in this parser were mainly stored in the lexicon.

A request could be in one of two states: active or inactive. A request was activated when the parser encountered a word whose dictionary entry contained that request. Once active, a request stayed active until it *fired*, or was executed; or until it was explicitly deactivated by another request. A request fired if it was in the active state and the conditions of active memory satisfied the test portion of the request's test-action pair.

Requests were largely responsible for building conceptual representations. Thus, common actions performed by requests were building an instantiation of a particular concept, or filling a slot in an already-built concept. For example, the dictionary entry of verb "ate" contained a request to build an instantiation of the concept INGEST (the Conceptual Dependency (Schank, 1972) primitive underlying eating and drinking); and a request to look for a noun group after the verb which referred to a food item, which, if found, was placed in the OBJECT slot of the INGEST.

Some requests in Riesbeck's parser were not stored in the lexicon. For example, at the beginning of a sentence a request was activated which looked for a noun group. When one was found, it was placed in a variable called #SUBJ. Later, when a request from a verb took the noun group from #SUBJ and placed it in the appropriate slot (usually the ACTOR slot) of the verb's representation. However, the number of requests like this was quite small, and thus most of the requests in the system were lexically-based.

The request-based method of parsing has been used in many other parsers since Riesbeck's parser. In the Conceptual Analyzer (CA) (Birnbaum and Selfridge, 1979), requests were used in much the same way as in Riesbeck's analyzer, but with the elimination of many non-lexically-based requests. Thus, instead of a request activated at the beginning of a sentence which looked for a noun group, CA had requests in the dictionary entries of verbs, looking back in the sentence for a noun group which could function as the subject. For instance, the verb "ate" had a request looking for a noun group to its left, which was an ANIMATE. If such a noun group was found, it was placed in the ACTOR slot of the INGEST. Other parsers using request-based knowledge include the Integrated Partial Parser (IPP) (Lebowitz, 1980), the Word Expert Parser (WEP) (Small, 1980), and BORIS (Dyer, 1982)[7].

[7] The test-action pairs in WEP and BORIS were called *demons*, rather than requests.

4.3 Integration and Request-based Parsing

One of the main tenets behind request-based conceptual analysis has been that the traditional separation of text analysis into morphological, syntactic, semantic, and pragmatic phases should be eliminated. This is for two reasons: first, given that the goal of conceptual parsing is to build a meaning representation, and not a syntactic analysis, there is no *a priori* reason for a separate syntactic analysis phase to exist. Second, since semantic and pragmatic information can sometimes help to eliminate syntactic, or even morphological, ambiguities, semantics should be brought into the parsing process as quickly as possible. The examples I presented in section 1.1 were examples of situations in which semantics/pragmatics must be used in order to eliminate syntactic ambiguities. Another such example was given in (Riesbeck and Schank, 1976), and involved the following sentence:

Hunting dogs can be dangerous.

Out of context, this sentence is syntactically (and semantically) ambiguous. However, in the following contexts, the ambiguity is eliminated:

Do you want to try shooting those dogs that have been pillaging the village? -- No, hunting dogs can be dangerous.

Let's take some dogs with us when we go to hunt moose. -- No, hunting dogs can be dangerous.

In the context of shooting dogs, in the first example, the words "hunting dogs" make more sense as a gerund and its object, since the semantic interpretation of this syntactic sense fits into the semantic context. However, in the second sentence, since the context is hunting moose, and since we know that dogs are often used to assist in the hunting of other animals, the better interpretation of "hunting dogs" is that "hunting" is an adjective, modifying "dogs."

The desire to do away with the separation of syntax and semantics has been articulated further in (Schank and Birnbaum, 1980) in terms of the *Integrated Processing Hypothesis*. Schank and Birnbaum suggest that the integration of syntax and semantics in a parser can be measured in terms of three aspects of the parser: its *control structure*, or the processes which occur during parsing; its *representational structures*, or the representations which it builds during the parsing process; and its *knowledge base*, or the set of rules which the parser draws on and which drive the parse of a text. A parser with an integrated control structure would have no separate syntactic or semantic processes or phases; one with integrated representational structures would not build any separate syntactic structure during the parsing process, but instead only semantic representations of its text; and a parser with an integrated knowledge base would have no rules which contained only syntactic knowledge.

The Integrated Processing Hypothesis states that syntax and semantics should be completely integrated in the control structure and representational structures, and that much of the knowledge base should also be integrated, although some separate syntax will exist in the knowledge base. Thus, it advocates that no separate phases of parsing should exist, that no purely syntactic information should be contained in the representations built during the parsing process, and that only some rules will exist in the parser's knowledge base which are purely syntactic.

In light of the goal of bringing semantic and conceptual information to bear as early as possible in the parsing process, at least some of the motivation behind the Integrated Processing Hypothesis is clear. Since semantics can aid even the earliest stages of the

parsing process, it is important that a parser's control structure be highly integrated. One way to ensure that the control structure is integrated is to integrate the semantic and syntactic knowledge in a parser as much as possible. Thus, its knowledge base should also be highly integrated. Finally, with regards to representational structure, the goal of conceptual parsers is to build a conceptual representation of the meaning of a text, not to produce a syntactic parse tree for the text. Therefore, syntactic representations should not be built during parsing unless it is clear that they aid in the process of building the conceptual representation. Given these motivations, then, the Integrated Processing Hypothesis takes almost as strong a stand on the integration of syntax and semantics as can be taken.

The result of these principles has been the use of lexically-based requests to encode syntactic knowledge. Lexically-based requests meet, more or less, with the criteria of the Integrated Processing Hypothesis. Certainly syntax and semantics are completely integrated in terms of process. Requests carry on in parallel any syntactic processing with the building of conceptual structures. Requests are also as integrated as possible with respect to the knowledge base, since they usually contain both syntactic and semantic knowledge. For example, the requests discussed earlier for the word "ate" contained the conceptual knowledge that "ate" refers to the concept INGEST, and that the ACTOR of INGEST should be an ANIMATE and the OBJECT of INGEST should be a food item. These same requests also contained the syntactic information that the ACTOR of "ate" occurs before the verb in an active sentence, and the OBJECT after the verb. In representational structures, also, the integration is high using requests, since no syntactic representations are explicitly built by the requests.

4.4 Problems With Integration of Parsing Knowledge

The motivation behind integration of syntax and semantics is well-grounded: it seems clear that semantics/pragmatics can sometimes assist during syntactic or morphological analysis. Thus, integrated processing is a desirable goal. However, the fulfillment of this goal via the complete integration of syntactic and semantic knowledge in terms of lexically-based requests leads to problems similar to the problems encountered with syntax-based transfer rules in chapter 2. This complete integration forces the encoding of all conceptual and syntactic knowledge at the same level of generality; namely, at the lexical level. All parsing knowledge in request-based parsers must be expressed at this level. It is not possible to encode rules which apply to syntactic classes of words, or to words which all mean the same thing. As we will see, this results in the need for many more rules than we would like.

4.4.1 Integrated Rules and Frame Selection

One issue which arises in conceptual analysis is due to the use of frames (Minsky, 1975) and other frame-like structures such as scripts (Schank and Abelson, 1977) to help in the parsing process, and to represent the meaning of the text. The *frame selection problem* (Charniak, 1982), or the selection of the appropriate frame for a text, must be faced by any conceptual analyzer using a large number of frames. How does a system choose the right frame from a large number of possible frames? Sometimes, particular words in a text point directly to a particular frame, thus trivializing this problem. For

example, the word "arrest" refers directly to a high-level structure, such as the $ARREST script. However, more often it is the case that no one word in a text points definitively to a unique frame. Instead, many of the words in the text are ambiguous or vague, and it is only by considering them in combination that a frame can be selected. An arrest, for instance, can be described without using the word "arrest," as in "Police took a suspect into custody," or even "They got their man." In cases like this, frame selection is much more difficult.

In request-based parsers, frame selection has usually been treated as a word disambiguation problem. In Riesbeck's parser, frame selection proceeded, by means of word disambiguation, in one of two ways, which more or less correspond to *bottom-up* and *top-down*. In the bottom-up method, the dictionary entry of a vague or ambiguous word which could refer to more than one frame (CD primitive) contained pointers (in the form of requests) to all the possible primitives to which it could refer. Thus, selecting a frame for a word was a matter of disambiguating the word to one of its meanings. A word was disambiguated when one of the requests in the dictionary definition of the ambiguous word fired, thereby choosing the word sense that it pointed to as the meaning of the word.

The bottom-up method performed the disambiguation of the word "wants" in the following two examples:

John wants Mary.
John wants the book.

The conceptual dependency parses for these two sentences are quite complicated, and the details of how "wants" is represented are not relevant here, since it is represented the same way for both sentences. What is represented differently in these two examples is the object of John's wanting:

```
                                                  --> John
JOHN <=> wants <-- MARY <=> PTRANS <-- Mary <--|
                                                  --< ?


                                                --> John
JOHN <=> wants <--  ? <=> ATRANS <-- book <--|
                                                --< ?
```

The difference between the representations reflects the fact that the first sentence means, more or less, the same as "John wants Mary to be near him," while the second sentence is closer to "John wants possession of the book to be passed to him."

In order to produce two different parses for these two sentences, Riesbeck's dictionary definition of the word "wants" contained two requests (among others), each of which would produce one of the two conceptual dependency configurations above. These requests, in prose form, were as follows:

> If "wants" is followed by a word which refers to an inanimate object, then the OBJECT of "wants" is an ATRANS of the inanimate object to the ACTOR of "wants."
>
> If "wants" is followed by a word which refers to a person, then the OBJECT of "wants" is a PTRANS of the person to the ACTOR of "wants."

At times, frame selection was also performed in a top-down fashion. This method was similar to the bottom-up method, in that frame selection was still treated as a disambiguation problem. In this method, however, the dictionary definition of the ambiguous or vague word consisted simply of a list of word senses. A request from some previous word in the text was then responsible for selecting a word sense, and thus selecting a frame to represent the ambiguous word.

The top-down method was used to select the CD primitive for the word "beat" in the following example:

John and Mary were racing. John beat Mary.

The dictionary definition of "beat" consisted of two senses, BEAT1 and BEAT2. BEAT1 corresponded to the "physical beating" sense "beat," while BEAT2 corresponded to the "victory" sense of "beat," as in the example above. BEAT1, the sense corresponding to a physical beating, was the default sense of the word. Thus, if no requests fired when the parser encountered the word "beat," it was taken to mean BEAT1. In the example above, however, the context of "racing" activated a request, which activated a "contextual cluster of conceptualizations." This cluster contained information about other conceptualizations which were likely to appear in a racing story, as well as information about which senses of ambiguous words would be used in a racing context. One piece of information in the cluster pointed to by "racing" was that the sense BEAT2, the sense meaning "victory," is the sense of "beat" used in racing stories. Thus, when the contextual cluster of conceptualizations was activated by the word "racing," a request was activated which expected the sense BEAT2 of "beat." When the parser encountered the word "beat," this request fired, and BEAT2 was activated instead of BEAT1.

Riesbeck's word disambiguation strategies were important in that they marked one of the first attempts at incorporating the conceptual context of a word into the process of its disambiguation. By using this method, many senses of a word could be eliminated which could not be eliminated by syntactic means. For instance, the disambiguation of the word "beat" in the example above could not be done on the basis of syntactic information, because there is not necessarily any syntactic difference between uses of the different senses of "beat."

Since the representations used in Riesbeck's parser were made up of the 10 or so Conceptual Dependency primitives, Riesbeck did not encounter the frame selection problems that arise in a system with a large number of frames. Since then, though, attempts have been made to apply this approach to frame selection in systems with larger number of frames, with mixed success. The BORIS system (Dyer, 1982) is an example of such a system. BORIS' representational vocabulary was much more diverse, and therefore there were many more frames to choose from in the system. The BORIS parser used lexically-based demons (requests) for frame selection in a top-down and bottom-up fashion, paralleling Riesbeck's two methods. For example, the word "gin" was disambiguated in a top-down fashion in the following two sentences:

John drinks gin.
John plays gin.

Demons stored as part of the dictionary definitions of "drinks" and "plays" performed the disambiguation of "gin" in these examples. "Drinks" loaded a demon which expected a liquid after the verb, while "plays" expected to find a game after the verb.

Bottom-up disambiguation in BORIS was slightly different from Riesbeck's bottom-

up method. Here, the parser tried to take advantage of the use of more high-level structures, by referring to these structures in the disambiguation demons. For the word "gin," bottom-up disambiguation rules were the following:

If the context is INGEST then interpret "gin" as a LIQUID.

If the context involves COMPETITIVE-ACTIVITY then interpret "gin" as a GAME.

Here, COMPETITIVE-ACTIVITY is a high-level structure in BORIS's representational vocabulary.

Unfortunately, rules such as this which refer to the general semantic context in which a word appears do not always work. For example, these rules would not disambiguate "gin" in the following example:

The gin spilled on the floor.

Here, since the context has nothing to do with either INGEST or COMPETITIVE-ACTIVITY, the parser would not be able to disambiguate "gin." Bottom-up context rules would have to mention all possible actions which liquids could be associated with in order to recognize the "liquid" sense of gin in all contexts. This list of actions could be quite long. Even with such a list, demons which looked for contexts could sometimes be misled:

The gin which the card players drank was bad.

In this case, the context includes both INGEST and COMPETITIVE-SITUATION, so again the disambiguation rules could not choose which sense of "gin" is appropriate.

One other parser which used a similar frame selection method was the Word Expert Parser (Small, 1980). Like BORIS, the the Word Expert Parser used demons to disambiguate words. WEP disambiguated very vague or ambiguous words, using complex dictionary definitions which consisted, in part, of a discrimination net of possible concepts to which an ambiguous word could refer, as well as a group of demons which were used to find the word's slot-fillers and to determine under what conditions the word referred to a particular concept.

An example of an ambiguous word which WEP disambiguated is the word "throw." Small considered several possible meanings of the word, such as "to throw out garbage", "to throw a party," "to throw in the towel," and "to throw a ball." Part of the dictionary definition of "throw" consisted of a discrimination net of the concepts to which "throw" could refer, such as PERSON-THROW, THROW-OBJECT-TO-LOCATION, THROW-OUT-GARBAGE, etc. Also included in the dictionary entry were demons to fill the slots of whatever concept "throw" referred to, and then determine which concept "throw" referred to based on the slot-fillings. Some of these demons were the following, in prose form:

Look for an active concept in memory and assign it as the agent of "throw."

If the agent of "throw" is a PERSON, then refine "throw" to PERSON-THROW.

Wait for a an concept after "throw" which is a MEAL, GARBAGE, a SMALL-PHYSOBJ, and PERSON, a CONTEST, or a PARTY, and assign it as the object of PERSON-THROW.

If the object of PERSON-THROW is GARBAGE, then refine PERSON-THROW to THROW-OUT-GARBAGE.

If the object of PERSON-THROW is a SMALL-PHYSOBJ, then refine PERSON-THROW to THROW-OBJECT-TO-LOCATION.

The rules which Small outlined were able to distinguish between many senses of the word "throw," but not without a price. The complexity and number of rules required to disambiguate "throw" was very high. This complexity, together with the flaws in BORIS' parsing rules, indicates that this approach to frame selection becomes more tenuous when dealing with a large number of frames. In fact, I will now argue that this approach suffers from the same sorts of problems as the syntax-based transfer rules from the last chapter.

Recall from earlier chapters the Spanish phrase "realizar diligencias" and the police investigation story:

Spanish: La policia REALIZA INTENSAS DILIGENCIAS PARA CAPTURAR a un reo.

English: The police ARE UNDERTAKING AN INTENSE INVESTIGATION in order to capture a criminal.

How could we write requests to disambiguate "realizar diligencias"? It would be difficult, if not impossible, to use this approach to frame selection for such a vague phrase as this. This technique would require a request in the dictionary definition of "realizar diligencias" for each possible meaning of the phrase. Thus, first we would need an exhaustive list of the possible representations which could be used for "realizar diligencias," so that we would know what requests would need to be written. For all practical purposes, this is an impossible task, since the phrase could conceivably refer to just about any action.

Even discounting this problem, though, writing lexically-based disambiguation rules for "realizar diligencias" would be a difficult task. Consider the requests which would be required just for the sense of "realizar diligencias" meaning POLICE-INVESTIGATION, as in the first example above. This is a similar task to writing transfer rules for this example. Upon first glance, one might think that it would be sufficient to check for the appropriate conceptualization, namely POLICE, appearing to the left of "realizar diligencias." In other words, whenever POLICE is the ACTOR of "realizar diligencias," the phrase means POLICE-INVESTIGATION. However, I gave a counterexample to this rule in chapter 2:

Spanish: La reina Isabela va a visitar a la ciudad de Nueva York el lunes. La policia REALIZA DILIGENCIAS para insurar su seguridad durante la visita.

English: Queen Elizabeth will visit New York city on Monday. The police ARE TAKING PRECAUTIONS to insure her safety during her visit.

Thus, requests must also check other portions of the sentence. These would have to be the same portions of the sentence that transfer rules had to check. Recall from before that this involved checking that the ACTOR of the "diligencias" was the POLICE, the "diligencias" were IN-SERVICE-OF a CAPTURE, and the OBJECT of the CAPTURE was a CRIMINAL. Requests would have to check for all of these features in the sentence. Thus, requests to disambiguate "realizar diligencias" in the above example would be the following:

REQUEST 1: If a word meaning POLICE appears to the left of "diligencias," then activate REQUEST 2.

REQUEST 2: If the preposition "para" (in order to) appears after "diligencias" and is followed by a word meaning CAPTURE, then activate request 3.

REQUEST 3: If a word meaning CRIMINAL appears after the CAPTURE, then fill the OBJECT slot of the CAPTURE with CRIMINAL and build the representation POLICE-INVESTIGATION for "diligencias." Fill the ACTOR slot of the POLICE-INVESTIGATION with POLICE.

These requests look very similar to the transfer rules for this example. Moreover, they have the same flaw: they are so tailored to this particular example that they will not work for semantically similar sentences which are worded differently. They will not work for this rewording:

Spanish: INTENSAS DILIGENCIAS POR PARTE DE LA POLICIA resultaron en la captura de un reo.

English: AN INTENSE POLICE INVESTIGATION resulted in the arrest of a criminal.

Just as for the transfer rules, a completely different set of requests would be required for this example, this time looking for "por" (by) followed by a word meaning POLICE appearing after "diligencias," then looking for the verb "resultar" (to result) followed by a word meaning CAPTURE, and finally looking for a word meaning CRIMINAL.

Judging from the number of requests needed to disambiguate "diligencias" in just these two examples, we can see that it is very difficult to use lexically-based disambiguation rules for very vague words like "diligencias" which have many possible meanings. First, the number of meanings of such words is very large. Second, even the number of rules for each possible meaning of a very vague word would have to be quite large, and each rule would have to be quite complex, due to the number of possible items in the surrounding context that could play a role in the disambiguation process.

This problem occurs with lexically-based requests for much the same reason that it occurred for syntactic transfer rules: the requests force us to encode knowledge at an inappropriate level of generalization. Since lexically-based requests completely integrate syntactic and conceptual knowledge, it is not possible to encode a frame selection rule which is based on conceptual knowledge without also including syntactic knowledge in the rule. What we really want is to encode rules like those I discussed in chapter 3, such as the Abstraction Inference Rule, and the Expected Action Inference Rule. Unfortunately, it is not possible to encode rules like these in request form, because requests always include syntactic information like "look to the left of the verb," etc.

4.4.2 Integration and Syntax

Similar rule proliferation problems occur in request-based parsers because syntactic knowledge cannot be expressed at the appropriate level of generality. Consider some of the requests which would be found in many conceptually-based parsers, such as CA. Under the word "gave" would be a request looking back in the sentence for the ACTOR of the ATRANS (the CD primitive representing "gave"). Similarly, under the verb "ate" would be a request looking back for the ACTOR of the INGEST (the CD primitive representing "ate"). These requests would have further restrictions on them as to where

the ACTORs could be found. These restrictions would reflect the fact that the subject of a sentence cannot in general appear in a prepositional phrase, or as the object of another verb, and so forth:

"Gave" request: Look back for a noun group which has the property ANIMATE, which is not the object of a preposition, or the object of a verb, or attached syntactically to anything before it. Place the conceptualization in the ACTOR slot of the ATRANS.

"Ate" request: Look back for a noun group which has the property ANIMATE, which is not the object of a preposition, or the object of a verb, or attached syntactically to anything before it. Place the conceptualization in the ACTOR slot of INGEST.

These requests are quite similar. They both involve filling the ACTOR slot of the conceptualization built by the verb with a noun group before the verb which is not syntactically attached to anything before it. All of this common information can be abstracted out, into a more general request, which could then be augmented by information from particular verbs:

ACTOR filling request: Look back for a noun group which has the property ANIMATE, which is not the object of a preposition, or the object of a verb, or attached syntactically to anything before it. Place this conceptualization in the ACTOR slot of the conceptualization built by the word which activated this request.

"Gave" information: "Gave" builds the conceptualization ATRANS.

"Ate" information: "Ate" builds the conceptualization INGEST.

Most other verbs in CA had similar requests, looking for a noun group before the verb, with the same syntactic restrictions on this noun group, to fill a particular slot in the conceptualization built by the verb. This slot is not always the ACTOR slot, as it is for "gave" and "ate," but there are still many similarities among these requests. Here are some examples:

"Received" request: Look back for a noun group which has the property ANIMATE, which is not attached syntactically to anything before it. Place the conceptualization in the RECIPIENT slot of the ATRANS built by "received."

"Talked" request: Look back for a noun group which has the property PERSON, which is not attached syntactically to anything before it. Place this conceptualization in the ACTOR slot of the MTRANS built by "talked."

Again, these two requests look for a noun group before the verb which is not attached syntactically to anything before it. There are some differences between these requests and the requests for "gave" and "ate." First, the "received" request fills the RECIPIENT slot instead of the ACTOR slot. Also, the "talked" request looks for a PERSON instead of an ANIMATE.

Again, the similarities among these four requests, and among similar requests found in the dictionary definitions of most verbs, can be abstracted out, to form a general

request that could apply to any verb. This request could be augmented, as before, wit[h] information from a particular verb. The general request would be the following:

Subject request: Look back for a noun group which is not attached syntactically to anything before it. This noun group fills a particular slot (ACTOR, by default) in the conceptualization built by the word which activated this request. The word which activated this request will provide the name of the slot which should be filled, if it is not the ACTOR slot. The conceptualization built by the activating word will provide semantic restrictions on the noun group to be chosen by this request.

Individual verbs, as well as the concepts built by these verbs, would provide th[e] information that was lost in the process of abstracting out the common information in th[e] original lexically-based requests:

"Gave" information: "Gave" builds the conceptualization ATRANS.

"Ate" information: "Ate" builds the conceptualization INGEST.

"Received" information: "Received" builds the conceptualization ATRANS. The slot to be filled by the subject request is RECIPIENT.

"Talked" information: "Talked" builds the conceptualization MTRANS.

ATRANS information: The ACTOR and RECIPIENT of an ATRANS are ANIMATE.

INGEST information: The ACTOR of an INGEST is ANIMATE.

MTRANS information: The ACTOR of an MTRANS is a PERSON.

The point of all this is to show the duplication of syntactic knowledge in request-based parsers. In CA, and in other request-based parsers, every verb required a reques[t] similar to the ones we have seen for "gave," "ate," "received," and "talked." Intuitively these requests all correspond to a single syntactic rule, having to do with where to find [a] verb's subject, and how to combine a verb and its subject semantically. In fact, if we tr[y] to abstract out the common information in these requests, we come up with a genera[l] request which corresponds to our intuitions. However, since request-based parsers forc[e] syntactic knowledge to be represented at the lexical level, and to be integrated completel[y] with semantic knowledge, the result is the need for a duplicate copy of nearly the sam[e] request in the dictionary entry of every single verb.

4.5 How Much Syntax is Necessary in Conceptual Analysis?

Thus far, I have argued that the complete integration of syntactic and semanti[c] knowledge in request-based parsers results in an inefficient encoding of this knowledge. [I] will now turn to another issue in conceptual analysis: the role of syntax in a conceptua[l] parser. By the nature of conceptual analysis, the role of syntax is different than in [a] syntactic analyzer. Since the final product of a conceptual analyzer is a conceptua[l] representation, syntax should play a role in parsing only if it helps to build the conceptua[l] representation.

In much of past conceptual parsing research, the assumption has been made that [a]

full-blown syntactic analysis is not needed in order to build a conceptual representation of text. Instead, many past conceptual analyzers have relied on "local" syntactic checks for the syntactic information needed.

To demonstrate this, let us examine some of the syntactic rules used in conceptual parsers. Consider the following sentence, which was parsed by CA:

A small plane stuffed with 1500 pounds of marijuana crashed.

The word "stuffed," as is the case with many English past participles, can function as either a past participle or a past active verb. In this context, it functions as a past participle, signaling the beginning of an unmarked passive relative clause.

In this sentence, CA built the following representation for the word "stuffed":

(PTRANS ACTOR ?ACTOR OBJECT ?OBJECT TO (INSIDE PART ?PART))

In this conceptualization, PTRANS is the conceptual dependency primitive for a change in physical location. The labels beginning with "?" indicate empty slots which need to be filled during the parse. Thus, to stuff something into a container is to PTRANS it to the inside of the container.

In order to parse this sentence, CA needed to determine whether the word "stuffed" functioned as a past active verb, a passive preceded by a form of "to be," or the beginning of an unmarked relative clause, and therefore passive. CA required three requests in the dictionary definition of "stuffed" to make this decision. One request looked for a form of the word "to be" before "stuffed." If it was found, then the ?PART position in the representation of "stuffed" was filled with the conceptualization to the left of the form of "to be." If this request did not fire, then a second request looked for the word "with" immediately following "stuffed," and expected to find the OBJECT being stuffed following "with." If this request fired, the verb was assumed to be the beginning of an unmarked clause, and again the conceptualization to the left of the verb was placed in the ?PART position in the above conceptualization. The firing of this request also resulted in the activation of another request which looked for another verb later in the sentence, indicating the end of the clause. The third request looked for an appropriate conceptualization following the verb which would fill the ?OBJECT position in the conceptualization. If this was found directly after the verb, then the verb was assumed to be active, and the conceptualization before the verb was placed in the ?ACTOR position.

In more precise terms, the requests required for this sentence were the following:

REQUEST 1:

TEST: A form of "to be" appears to the left of "stuffed."

ACTION: Fill the ?PART position of the conceptualization built by "stuffed" with the conceptualization to the left of the form of "to be," and deactivate requests 2 and 3.

REQUEST 2:

TEST: The word "with" appears after "stuffed."

ACTION: Fill the ?PART position of the conceptualization built by "stuffed" with the conceptualization to the left of "stuffed," remember all the conceptualizations in active memory to the left of "stuffed," load REQUEST 2A, and deactivate request 3.

REQUEST 2A:

TEST: A verb has been found.

ACTION: Reset active memory to the state remembered in the ACTION of REQUEST 2.

REQUEST 3:

TEST: A conceptualization which can function as a container has been found after "stuffed."

ACTION: Fill the ?PART position of the conceptualization built by "stuffed" with the container conceptualization, fill the OBJECT slot with the conceptualization to the left of "stuffed," and deactivate request 2.

These requests used "local" syntactic information in order to disambiguate the word "stuffed." By this, I mean that only words in the immediate neighborhood of "stuffed" were checked for particular syntactic properties or for their presence or absence. If a form of "to be" appeared directly before "stuffed," then "stuffed" was assumed to be passive, but not part of a relative clause. If the preposition "with" appeared directly after "stuffed," then "stuffed" was part of an unmarked relative clause. If a noun group followed "stuffed" which could function as its direct object, then "stuffed" was a past active verb.

The advantage of using only local syntactic checks in requests was that it was not necessary to keep track of the syntactic "state" of the parser. Rather than having to rely on rules like "the main verb of the sentence has not been found yet, so 'stuffed' might be a past active", or "the main verb of the sentence has already appeared, so 'stuffed' must be an unmarked passive", which would require that the parser keep track of syntactic states like "the main verb has been found," using only local syntactic properties allowed the parser to get away with less syntactic bookkeeping. Thus, parsers like CA tried to get away with using only local syntactic checks in their requests.

However, it is not always the case that local checks like those used to disambiguate "stuffed" are enough. Consider the following examples:

The soldier called to his sergeant.
I saw the soldier called to his sergeant.

The slave boy traded for a sack of grain.
I saw the slave boy traded for a sack of grain.

In these cases, the appearance of a preposition after the verbs "called" and "traded" does not guarantee that the verbs are passive. This is because both verbs can be used

either transitively or intransitively. Instead, the information that must be used to determine whether the verbs are active or passive is whether or not there is another verb in the sentence which functions as the main verb.

Writing requests to disambiguate "called" or "traded" using only "local" syntactic checks would not be as easy as for "stuffed." First, words such as "called" would require a request looking to the left to see if another verb is already on the active list. If a verb is found, then "called" must be unmarked passive. However, if no verb is on the active list, this does not guarantee that "called" is active, since the main verb of the sentence could also come after "called," as in the following example:

The soldier called to his sergeant was reprimanded.

Thus, two requests would be required, one looking back for the main verb of the sentence, and one looking forward for the main verb.

The requests needed to determine whether verbs such as "called" are active or passive, then, would be the following:

REQUEST 1:

TEST: A form of "to be" appears to the left of "called."

ACTION: Fill the OBJECT slot of the MTRANS built by "called" with the conceptualization to the left of the form of "to be," and deactivate requests 2-4.

REQUEST 2:

TEST: A verb is in active memory to the left of "called."

ACTION: Fill the OBJECT slot of the MTRANS built by "called" with the conceptualization to the left of the verb, and deactivate requests 1, 3, and 4.

REQUEST 3:

TEST: An active verb has been found to the right of "called."

ACTION: Fill the OBJECT slot of the MTRANS built by "called" with the conceptualization to the left of the verb, and deactivate requests 1, 2, and 4.

REQUEST 4:

TEST: The end of the sentence has been found, and no active verbs are to the left or the right of "called."

ACTION: Fill the ACTOR slot of the MTRANS built by "called" with the conceptualization to the left of the verb, and fill the RECIPIENT slot of the MTRANS with the conceptualization after the verb, or after the word "to."

There is an additional problem with performing the resolution of ambiguous verbs like "called" with local syntactic checks. That is the interaction between two such ambiguous verbs in the same sentence. Consider the following examples:

The soldier called to the sergeant shot in the arm.
The soldier called to the sergeant shot three enemy troops.

As these examples illustrate, it is not enough to look for a verb further on in the sentence, because that verb may also be either past active or past participle.

To handle examples like these, the requests above would have to be made still more complicated. A request under "called" would have to look for a verb which could either be past active or past participle. If such a verb was found, then special requests would have to be activated which would look for the appropriate clues around the second verb to determine whether it was active or passive, thus also determining if the first verb was active or passive.

These additional requests would be the following:

REQUEST 5:

TEST: A verb appears after "called" which could either be past active or past participle.

ACTION: Activate special requests for that verb which determine whether that verb is past active or past participle.

SPECIAL REQUESTS FOR "SHOT":
REQUEST 6:

TEST: A form of "to be" appears to the left of "shot." (indicating that "called" was an unmarked passive, as in "The soldier called to his sergeant was shot")

ACTION: Fill the OBJECT slot of the conceptualization built by "shot" with the conceptualization to the left of the form of "to be," fill the OBJECT position of the MTRANS built by "called" with the conceptualization to its left, and deactivate requests 7 and 8.

REQUEST 7:

TEST: The word "in" appears after "shot." (indicating that "called" was the main verb of the sentence, as in "The soldier called to the sergeant shot in the arm.")

ACTION: Fill the OBJECT slot of the conceptualization built by "shot" with the conceptualization to its left, fill the ACTOR slot of the MTRANS built by "called" with the conceptualization to its left, and and deactivate request 8.

REQUEST 8:

> TEST: A conceptualization which is a PHYSICAL-OBJECT has been found after "shot." (indicating that "shot" is the main verb of the sentence, and "called" was an unmarked passive, as in "The soldier called to the sergeant shot three enemy troops.]"
>
> ACTION: Fill the OBJECT slot of the conceptualization built by "shot" with the PHYSICAL OBJECT, fill the ACTOR slot with the conceptualization to the left of "shot," fill the OBJECT slot of the MTRANS built by "called" with the conceptualization to its left, and deactivate request 7.

There are still examples for which even this complex set of requests would not be enough:

> The soldier called to the sergeant shot in the arm was reprimanded.

In this sentence, even though "shot" is part of a relative clause, "called" is still not the main clause verb, since "was reprimanded" follows later in the sentence. Thus, the above requests will fail to parse this sentence correctly.

Out of context, the last example is not an easy one to understand. Thus, one might argue that it is not necessarily a bad thing that the above requests would not be able to parse the sentence. However, there are contexts in which this construction would be quite natural, as in:

> Several soldiers got drunk in their barracks and shot up their boot camp, shooting one sergeant in the arm. After the incident, each soldier was called to his officer, to be appropriately disciplined. The soldier called to the sergeant shot in the arm was severely reprimanded.

Given that this sentence can be easily understood in the appropriate context, it is important to be able to write rules which can correctly parse it.

Thus, to take care of the ambiguities of English words which can function as either past actives or past participles, we see that local syntactic checks do not suffice. Although some of these words can use a set of requests which determine from the local context whether they are active or passive, such as the set of requests for "stuffed" above, this approach cannot work for other verbs, such as "called" and "traded." These verbs require more requests, which look for the presence of other past active verbs in the sentence. These requests are quite complicated, because the verbs which they are looking for can be ambiguous themselves. Even with this increased complexity, examples can still be found which the requests do not cover.

So it appears that the syntactic ambiguity that many English verbs have between past active and past participle cannot be solved without great difficulty by local syntactic checks. This is because determining whether a verb is passive or active sometimes requires knowing whether or not another verb is functioning as the main clause verb. English sentences have only one main verb, and so when we encounter a verb which could be a past participle, we can use this fact to help us. Intuitively, we would like rules which say, "If there is no other main verb in the sentence, then the ambiguous verb must be past active, but if there is another main verb, then it must be past participle." However, we cannot formulate the rule in this fashion using local syntactic checks, because the knowledge as to whether or not another verb in the sentence is functioning as the main

clause verb requires more than just local syntax-checking.

4.6 Lexically-based Requests in a Multi-lingual Parser

In a multi-lingual machine translation environment, it would be convenient to use the same parser for each source language in the system. In a multi-lingual parser, it is desirable, for pragmatic reasons if nothing else, to share knowledge between languages whenever possible. Doing this makes the addition of more languages to the system easier, since the amount of new knowledge for each language is smaller. Also, unless this is done, it is not clear what it means to have a multi-lingual parser. If no rules in the parser are shared across languages, then the multi-lingual aspect of the parser is rather nebulous. One might just as well write a separate parser for each language.

It is not possible to share very much knowledge between languages using lexically-based syntax rules. This is because *it is not possible to ferret out the commonalities between languages from lexically-based rules*. Thus, a multi-lingual parser using lexically-based rules must have a great deal of its syntactic and conceptual knowledge duplicated between languages, making the addition of new languages harder in such a system.

Consider the lexically-based requests which would be found in the dictionary definition of the English word "shot." They would include requests looking left for the actor of the shooting, and looking right for the object; possibly a request looking for the preposition "in" after "shot" which would assign the object of "in" to fill a particular slot of the conceptualization built by "shot"; requests to determine whether "shot" is a past active or past participle; and the special requests which I described above for determining whether other verbs in the sentence were past active or past participle. Here is a list of these requests:

REQUESTS FOR DETERMINING WHETHER "SHOT" IS PAST ACTIVE OR PAST PARTICIPLE

REQUEST 1:

TEST: A form of "to be" appears to the left of "shot."

ACTION: Fill the OBJECT slot of the conceptualization built by "shot" with the conceptualization to the left of the form of "to be," and deactivate requests 2-4.

REQUEST 2:

TEST: A verb is in active memory to the left of "shot."

ACTION: Fill the OBJECT slot of the conceptualization built by "shot" with the conceptualization to the left of the verb, and deactivate requests 1, 3, and 4.

REQUEST 3:

TEST: An active verb has been found to the right of "shot."

ACTION: Fill the OBJECT slot of the conceptualization built by "shot" with the conceptualization to the left of the verb, and deactivate requests 1, 2, and 4.

REQUEST 4:

TEST: The end of the sentence has been found, and no active verbs are to the left or the right of "shot."

ACTION: Fill the ACTOR slot of the conceptualization built by "shot" with the conceptualization to the left of the verb, and fill the RECIPIENT slot of the conceptualization with the conceptualization after the verb, or after the word "to."

REQUEST 5:

TEST: The word "in" follows the object of "shot," and a conceptualization which is a BODYPART follows the word "in."

ACTION: Fill the HURT-PART slot of the conceptualization built by "shot" with the BODYPART to the right of "in."

REQUEST 6:

TEST: A verb appears after "shot" which could either be past active or past participle.

ACTION: Activate special requests for that verb which determine whether that verb is past active or past participle.

SPECIAL REQUESTS FOR "shot", ACTIVATED BY OTHER PAST ACTIVE / PAST PARTICIPLE VERBS:

REQUEST 7:

TEST: A form of "to be" appears to the left of "shot."

ACTION: Fill the OBJECT slot of the conceptualization built by "shot" with the conceptualization to the left of the form of "to be," fill the OBJECT position of the MTRANS built by "called" with the conceptualization to its left, and deactivate requests 7 and 8.

REQUEST 8:

TEST: The word "in" appears after "shot."

ACTION: Fill the OBJECT slot of the conceptualization built by "shot" with the conceptualization to its left, fill the ACTOR slot of the MTRANS built by "called" with the conceptualization to its left, and and deactivate request 8.

REQUEST 9:

TEST: A conceptualization which is a PHYSICAL-OBJECT has been found after "shot." (indicating that "shot" is the main verb of the sentence, and "called" was an unmarked passive, as in "The soldier called to the sergeant shot three enemy troops.")

ACTION: Fill the OBJECT slot of the conceptualization built by "shot" with the PHYSICAL OBJECT, fill the ACTOR slot with the conceptualization to the left of "shot," fill the OBJECT slot of the MTRANS built by "called" with the conceptualization to its left, and deactivate request 7.

Given a dictionary entry for "shot" like this, say we would like to define the Spanish word for "shot," "fusilar." Is the English dictionary definition any help to us? We would like it to be. "Fusilar" builds the same conceptualization as "shot"; it has the same slot-filling properties (the ACTOR still normally comes to the left, and the OBJECT to the right, of the verb); the same semantic restrictions on its slot-fillers apply; "en," the equivalent Spanish preposition to "in," can also be used after "fusilar" to refer to the part of the body injured in the shooting; etc.

However, given the organization of syntactic knowledge using lexically-based requests, it is not clear how any of these commonalities between the Spanish and English verbs for "to shoot" will help in the writing of the Spanish verb's dictionary entry. First, the English verb, "shot," can be either past active or past participle. Because of this, most of the slot-filling requests for ACTOR, OBJECT, and HURT-PART are the same requests which determine whether "shot" is past active or past participle. This syntactic ambiguity does not occur in Spanish. Therefore, it is not clear that the Spanish verb could use *any* of the same requests as the English verb. Perhaps request 5, which looks for the preposition "in," could be used with minimal modification. But the other requests are completely useless to us in building the dictionary entry for "fusilar." Thus, all the syntactic and conceptual knowledge in the requests of "shot" would need to be duplicated in the requests of "fusilar."

In lexically-based syntax rules, knowledge about different classes of words is inextricably intertwined together. In this case, knowledge specific to the word "shot," namely that it builds a particular conceptualization, and that certain slots of this conceptualization can be found in certain syntactic positions, is intertwined with other information, such as how to decide if "shot" is a past active or a past participle. This latter information is not specific to "shot." It is knowledge that is common to all past active / past participle verbs in English. Because this knowledge is not separated out in lexically-based requests, knowledge common to other languages cannot be shared. In Spanish, and in many other languages, there is a word corresponding to the English "to shoot." However, the information that these words share in common cannot be shared between dictionary definitions of the words, due to the fact that other information which these words do not have in common, such as how to tell a past active from a past participle, is inextricably intertwined with the shared information.

What we would like, then, is to be able to express syntactic rules in such a way that information which is shared among words in different languages is reflected in similarities in the dictionary definitions of these words in the parser. We would also like to be able to express differences between languages *in the simplest way possible*. For instance, the English "to look for" is equivalent to the Spanish verb "buscar," except the Spanish verb

does not take a preposition after it, as does "look." We would like the dictionary definition of "buscar" to reflect this difference in the simplest way possible, so the rule expresses the fact that "buscar" is *just like* "to look for," except the OBJECT being looked for is expressed as the direct object of "buscar," instead of the object of the preposition "for."

4.7 Conclusion

I have examined request-based parsers, and found them to be lacking in some respects with regards to the problems of machine translation. First, the use of lexically-based requests forces parsing knowledge to be encoded at the lexical level. This level of generality is inappropriate for much of the parsing knowledge we would like to include in the system. In the last chapter, the disambiguation knowledge which was needed for the translation of ambiguous or vague words was not easily encodable in terms of syntactic transfer rules. Unfortunately, the same is true with lexically-based requests, because the frame selection knowledge needed in a conceptual analyzer is likewise not appropriately encoded at this level of generality. For the police investigation example, we would like to encode a frame selection rule like "If police are the ACTORs of an action which is in service of the capture of a criminal, then the action is most likely a POLICE-INVESTIGATION." However, lexically-based requests do not allow us to encode a rule like this, because syntactic information must also be included in these requests.

A similar observation about lexically-based disambiguation rules was made in (Schank, Birnbaum, and Mey, 1983). Schank *et. al.* noted that vague words like "take," "use," etc., would require an explosively large number of distinct word senses. They asserted that this problem arises because the word sense disambiguation approach to frame selection "remains, at root, based on the old notion that the meaning of an utterance is a simple, additive function of the meanings of the words it contains." Schank *et. al.* did not propose a solution to the problem of large numbers of word senses, except to say that the definitions of vague words should consist of "crude descriptions" of what they might mean in a given context. Then, this crude description would be used as a "search key" for indexing inside of more specific frames, to try and find a match between the crude description and a more specific frame.

Similar problems are encountered with the encoding of syntactic information in the form of requests. It would be appealing to be able to encode syntactic knowledge like "The noun group to the left of a verb fills a slot (ACTOR, by default" of the conceptualization built by the verb.) However, using lexically-based requests, we can only encode rules for particular verbs, such as "A noun group to the left of the verb "ate" fills the ACTOR slot of the concept INGEST built by "ate."" Similar rules must be duplicated in the dictionary entries of every single verb, thus resulting in a much larger number of requests in the system than we would like.

Previous work in conceptual analysis is also lacking in other respects. For complex syntactic constructions, it is sometimes difficult to construct the requests that would be required for the disambiguation of these constructions. This is because requests rely on "local" syntactic checks, rather than using more global syntactic information which would require keeping track of the syntactic state of the parser. For verbs which can function as past participles or past actives, we would like to write rules like "If the main verb of the sentence has already been found, then the verb in question is a past participle." However,

this sort of rule is not possible using only local syntactic checks, because syntactic information like "the main verb of the sentence has already been found" is not computed. Thus, the requests to handle past active / past participle verbs are quite complex, and it is difficult to write requests that work in all cases.

In a multi-lingual environment, such as is required in machine translation, requests do not facilitate the sharing of common knowledge across languages. Since syntactic information, such as how to distinguish between a past active and past participle in English, in mixed in with other knowledge which might have more relevance to other languages, this other knowledge is not in a form that makes it easily applicable to other languages in the parser.

An argument can also be made against lexically-based parsing knowledge with regards to learning. This is true with regards to conceptual and syntactic knowledge. Consider the case of conceptual knowledge encoded in a lexically-based form. The fact that it is in the lexicon implies that this knowledge is strictly linguistic, and once learned, cannot be applied in other domains of human thought. But this is clearly not what we would like. Consider some of the rules found in the dictionary definition of "throw" in the Word Expert Parser:

> *If the agent of "throw" is a person, then refine "throw"* to PERSON-THROW.
>
> If the object of PERSON-THROW is garbage, then refine PERSON-THROW to THROW-OUT-GARBAGE.

Since these rules are lexically-based, this implies that they would not be available to non-linguistic inference processes. But we would want the same knowledge available to a vision system, for instance, observing someone throwing out garbage. If a vision module identified that a person was the agent of the action of throwing something, and if it also identified the object being thrown as garbage, we would want this system to also be able to make the inference that the garbage was being disposed of, or thrown out, not just that the garbage was being transported from one place to another by means of throwing it. This inference process is very similar to the process of disambiguating the word "throw" with the above rules. Yet the fact that the disambiguation rules are lexically-based implies that if the parser learned these rules, they would not be available to a vision module, or vice versa. Thus, we would want the knowledge used by these two processes to be shared between them.

An argument against lexically-based knowledge can also be made with regards to the learning of syntactic knowledge. If the syntactic rules in a parser do not reflect the generalizations that can be made about different classes of words in a language, it is difficult to imagine how the learning of a new word would proceed. For instance, consider the requests which I presented above for the word "shot." Some of these requests, such as the request looking for the preposition "in" after the verb, are rather specific to the verb "shot," and do not apply to other verbs. However, other requests, such as those which determine whether "shot" is past active or past participle, could apply, with a small amount of modification, to a larger class of verbs, namely those verbs which can be either past actives or past participles. Finally, other information in the requests, such as the fact that when "shot" is active, the ACTOR of "shot" often appears to the left of the verb, and the OBJECT to the right, applies to verbs in general. However, nowhere in these requests is this stated. All of the requests are written specifically for the verb "shot."

The learning problem, then, is that when a new verb is learned, the learner cannot

determine which requests that he knows from other verbs can apply to the new verb. Is it the case for the new verb that the preposition "in" will be followed by the HURT-PART slot of its conceptualization? Or should the learner infer that the new verb builds the same conceptualization as "shot"? How about the rules which determine whether "shot" is past active or unmarked passive? Do these rules apply to the new verb? In short, since none of this knowledge is marked as to how general it is, a learner cannot infer whether or not any of it applies to a new verb just being learned. Since this is the case, this implies that the learner would have to learn *everything* about how this new verb functions, including where to look in the sentence for the slot-fillers of its conceptualization, how to disambiguate it if it is ambiguous, what particular prepositions indicate particular slots, etc.

Clearly if nothing can be inferred about a new word from words that are already known, the task of learning an entire language would be hopelessly complex. A learner would be forced to learn an entirely new and intricate set of rules for every single word in the language. This is a ridiculously hopeless task, given the number of words in natural languages. So the lexically-based approach to syntactic knowledge appears to be incompatible with the task of learning a natural language.

I have confined the discussion in this chapter to request-based parsers, but many of the criticisms also apply to other previous integrated parsers. An example is Wilks' parser, (Wilks, 1973) part of his English-to-French machine translation system. This parser was integrated, in that any syntactic processing took place at the same time as semantic processing. It also shared many other of the properties of integrated parsers which I outlined in chapter 1.

Wilks' parser contained three types of structures to encode syntactic and semantic information: *elements*, *templates*, and *formulas*. *Elements* were a collection of 60 primitives, consisting of "entities" such as MAN and THING; "actions" such as FORCE, CAUSE, and FLOW; etc. Elements were the building blocks of *formulas*, which expressed the various senses, or meanings, of words. The meaning of "drink," for example, was expressed by the following formula:

((*ANI SUBJ) (((FLOW STUFF) OBJE) ((*ANI IN) (((THIS (*ANI (THRU PART))) TO) (BE CAUSE)))))

This formula encoded that "drink" is an action, done by animate things (*ANI SUBJ) to liquids ((FLOW STUFF) OBJE), causing the liquid to be in the animate thing (*ANI IN).

Templates consisted of strings of formulas, which were meant to encode common messages which were conveyed in natural language. One such template consisted of the sequence MAN-FORCE-MAN, meaning that one message that text can convey is that one man does something to force another man to perform an action.

The parsing process consisted largely of attempting to match natural language input to templates. When the sequences corresponding to known templates were found, then the sentence was parsed. This template-matching process performed word disambiguation, by eliminating possible formulas that corresponded to possible senses of a word which did not match possible templates. Thus, for example, in the sentence "Small men sometimes father big sons," "father" could be interpreted as a noun meaning MAN or a verb meaning "to cause to have life." These two interpretations would result in the following sequences of formulas:

KIND MAN HOW MAN KIND MAN
KIND MAN HOW CAUSE KIND MAN

Since the second interpretation matched a known template in Wilks' system, (MAN CAUSE MAN), this interpretation was chosen, and thus "father" was disambiguated to mean "to cause to have life."

Linguistic information in Wilks' system was implicitly encoded in the parser's templates. The order of the arguments in these templates reflected the nature of English syntax. Thus, the actions (FORCE CAUSE, etc.) in the templates appeared as the second formula, reflecting the fact that English is an SVO language.

Because of the mixing of syntactic and semantic information in templates, Wilks' system is subject to many of the criticisms that I have made about request-based systems. First, as with requests, semantic information cannot be encoded in the form of templates without implicitly including syntactic information. Thus, the number of templates that would be required for disambiguation of vague words would be unnecessarily large. This is because a given piece of semantic information which might determine the meaning of a word would have to be duplicated in templates corresponding to all of the syntactic constructions that this information could be conveyed in. For example, just as many requests were required to look for POLICE, CAPTURE, and CRIMINAL in the "diligencias" example, depending on where in the sentence these concepts appeared, many different templates would have to be written, corresponding to the different possible syntactic constructions that could be used with "diligencias" to mean "investigation."

Another difficulty with templates that was also true of requests is that syntactic information that intuitively seems like a single syntactic rule must be duplicated many times in different templates, rather than be expressed in terms of one rule. For example, information that the subject of a verb comes before the verb in English is implicitly encoded into every template that has an action as its second argument. Just as the request looking for an ANIMATE to the left of the verb "ate" encoded the syntactic information that the subject of an English verb appears to its left, the template (MAN CAUSE MAN) encodes the same information by virtue of MAN appearing to the left of CAUSE in the template. And just as this information had to be duplicated in the dictionary entry of every verb, this information is duplicated in every template in Wilks' system.

The use of templates also prohibits the use of the same semantic information to parse other languages. For example, in German, we would want to be able to use the information that (MAN CAUSE MAN) is a reasonable message. However, since German direct objects sometimes precede the verb rather than follow it, this template will not always match German texts which convey this message. We would need another template, (MAN MAN CAUSE), for these cases. Thus, conceptual information in Wilks' system cannot be shared between languages without significant modification.

Instead of expressing parsing knowledge in an integrated form, as is the case with requests and templates, what we would like instead is to express parsing knowledge at its correct level of generality, whatever that is. Certainly some information is specific to particular words, and would be suitably encoded as lexically-based requests. For instance, we know that the preposition "for" will precede the object of the verb "to search." This appears to be an ungeneralizable piece of syntactic knowledge, which only applies to the verb "search." However, a great deal of the conceptual and syntactic knowledge that I have examined thus far can be expressed in more general terms. In the police investigation story, general conceptual rules, such as the rules which were discussed in

chapter 3, seemed appropriate. Likewise, much of the syntactic knowledge in lexically-based requests can be generalized to larger classes of words, such as past participle verbs, verbs which can be either past participle or past active, or even to all verbs. This knowledge should be expressed at the right level of generality.

Thus, it appears that we should weaken the criteria of the Integrated Processing Hypothesis. Keeping in mind that the control structure of a parser should remain integrated, so as to facilitate the integration of syntactic and semantic processing, the representational structure and knowledge base should be made less integrated, so as to facilitate the ability to represent conceptual and syntactic knowledge in the parser at the right level of generality. In the next chapters, I will examine the form which this generalized conceptual and syntactic knowledge should take.

5. Using Hierarchical Memory Organization in Frame Selection

5.1 Introduction

In the last chapter, I argued that past work on conceptually-based word disambiguation suffers from some of the same problems as syntax-based techniques. Parsing knowledge is inefficiently encoded, because it is all lexically-based. Also, since conceptual and syntactic knowledge are completely integrated, it is not possible to express purely conceptual knowledge or purely syntactic knowledge. As a result, disambiguation rules must always depend on the word order of the surrounding context. Thus, semantically similar sentences with different syntactic constructions often require separate disambiguation rules, even though intuitively a single conceptual fact should do. Similarly, since purely syntactic knowledge must be mixed with conceptual information, syntactic rules must be duplicated in the lexical entries of many words of the same syntactic class, instead of having a single syntactic rule which governs the class.

In the next two chapters, I will present a more autonomous approach to the representation of conceptual and syntactic knowledge in a parser. This approach is used in the MOPTRANS parser, and overcomes some of the difficulties which I discussed in the last chapter, by allowing knowledge to be represented at different levels of generality. Purely conceptual or purely syntactic knowledge can be represented as such, while knowledge which is dependent on both syntax and semantics still can remain integrated.

Although the parsing knowledge in the MOPTRANS parser is less integrated than in previous conceptual analyzers, the parsing process is still very much integrated, in that syntactic and semantic/pragmatic processing of an input text occurs in parallel. The advantages of an integrated processing model of parsing are preserved, due to the way in which the parser utilizes the more autonomous syntactic and conceptual knowledge. This will be discussed in greater detail in chapter 6.

This chapter will be devoted to discussing the organization of conceptual knowledge in MOPTRANS. Instead of using lexically-based requests to encode conceptual parsing knowledge, MOPTRANS uses a small number of frame selection rules, similar to the rules which I sketched out in chapter 3; in conjunction with a hierarchically-organized conceptual knowledge base. The MOPTRANS parser is able to use a small number of frame selection rules because the conceptual knowledge in the parser is represented at the appropriate level of generality.

5.2 Frame Theory and Levels of Generality

Similar problems of duplication of knowledge have been encountered in frame-based systems before. It is worthwhile to examine the solutions that have been proposed, because these solutions have some bearing on the way in which the duplication of knowledge in syntax-based transfer rules and in request-based disambiguation rules can be

eliminated.

Charniak (Charniak, 1977) observed that, for efficiency reasons, causal knowledge in frames should not be duplicated when that knowledge comes from more general causal laws. Thus, in his representational system, which represented knowledge about mundane painting, he distinguished between two types of frames: *simple events*, which corresponded to common sense causal laws; and *complex events*, which referred to the simple events for their causal explanations. The frame PAINTING was a complex event, which consisted of sequences of actions such as "get paint on the painting instrument," and "bring instrument in contact with object", along with pointers to simple events, like STICK (i.e., "sticks to"), which provided the causal rules explaining why events within the PAINTING frame proceeded in that order. STICK consisted of a causal rule, explaining why bringing the painting instrument in contact with the object to be painted would cause paint to stick on the object. These simple events, like STICK, could be shared between many complex events, like PAINTING, so that knowledge common to more than one situation would not have to be duplicated in the frames used to represent those situations.

For different reasons, a similar sharing of knowledge was proposed by Schank in (Schank, 1982). This sharing of knowledge was a modification of script theory (Schank and Abelson, 1977). Scripts were originally proposed, and first used, as processing structures to help understand situations in which a standard set of actions usually occur, such as in a restaurant. Scripts were intended to facilitate inferencing about these situations, such as that in a restaurant paying comes after the meal (see (Cullingford, 1978)).

Schank discussed in (Schank, 1982) the use of scripts for learning. This new task raised some problems with scripts. In script theory, although similar scenes appeared in different scripts, the representation of these scenes did not capture these similarities. For instance, the scripts $RESTAURANT and $DEPARTMENT-STORE both contained a scene having to do with ordering. In the case of $RESTAURANT, this scene was ORDER-FOOD, and in $DEPARTMENT-STORE, the scene was ORDER-MERCHANDISE. However, these two scenes were totally separate entities, with no representation of the fact that they were similar in very important ways. There was no more general concept ORDER to which they could refer which was an abstraction of the common entities of both scenes. This presented a problem for learning. If a program were to learn scripts like $RESTAURANT and $DEPARTMENT-STORE, knowledge common to the ordering in a restaurant and the ordering in a department store would have to be stored in two different places, since the corresponding scenes in $RESTAURANT and $DEPARTMENT-STORE did not share information. This meant that knowledge learned in one domain could not be accessed in the other domain. So, for example, if one learned that one should be polite in order to get good service in a restaurant, that information could not be accessed by $DEPARTMENT-STORE in order to realize that one should be polite to the catalog clerks in order to get good service at a department store, also.

There was another problem with scripts, etc., which became evident from an experiment presented in (Bower, Black and Turner, 1979). In this experiment, recognition confusions were found to occur between stories about visits to the dentist and visits to the doctor. Intuitively, this result was not surprising, since most people have experienced such confusions. But how could such confusions be explained by scripts? Should we posit a "visit to a health care professional" script to explain it? Clearly, this would be beyond the initial conception of what a script was.

To accommodate solutions to these problems, a modification of script theory was proposed in (Schank, 1982) that introduced a new processing structure, called a MOP (Memory Organization Packet). The general idea behind MOPs was to store knowledge which is common to many different situations in only one processing structure, and then to make this processing structure available in all the different situations in which it applies.

To see how MOPs differ from scripts, let us compare the script and MOP representations of several events as presented in Figure 5-6. In script theory, all the scenes of the doctor, lawyer, or car wash episode were provided by one structure, $DOCTOR, $LAWYER, or $CAR-WASH. However, although similar scripts had many similar scenes, such as HAVE-MEDICAL-PROBLEM and HAVE-LEGAL-PROBLEM in $DOCTOR and $LAWYER, there was no connection between such scenes. In MOP theory, however, all the common elements shared among specific scenes of different contexts are abstracted together into a more general scene. Thus, the common features of doctor visits and lawyer visits are abstracted into a more general structure, M-PROFESSIONAL-OFFICE-VISIT (or M-POV for short). M-POV has scenes, WAITING-ROOM, GET-SERVICE, etc., which are abstractions of the scenes DOCTOR-WAITING-ROOM, LAWYER-WAITING-ROOM, and GET-TREATMENT, LEGAL-CONSULTATION. Then, features unique to doctors' offices are provided by the more specific scene DOCTOR-WAITING-ROOM, but features shared by other professional office visits exist in the generalized scene WAITING-ROOM. Similarly, the sequential information shared by these scripts is abstracted together also into M-POV. Thus, a great deal of information that was in $DOCTOR in script theory is not in M-DOCTOR in MOP theory. Rather, much of this information comes from more general MOPs.

Similarly, information which is shared between professional office visits and other types of getting service, such as getting your car washed, is abstracted even higher, into M-GET-SERVICE. All get-service actions have things in common, such as first having a problem, waiting for the service, and paying. This common information is abstracted to as general a structure as possible, into even more general scenes such as NEED-SERVICE and WAIT.

MOPs and scenes, then, are arranged hierarchically. M-POV is a generalization of the common elements in M-DOCTOR and M-LAWYER, M-GET-SERVICE is a generalization of M-POV and other types of service MOPs, WAITING ROOM is a generalization of DOCTOR-WAITING-ROOM and LAWYER-WAITING-ROOM, WAIT is a generalization of WAITING-ROOM and WAIT-IN-LINE, etc. Knowledge is stored at as general a level as is possible.

5.3 Using Hierarchical Memory Organization for Frame Selection

The sharing of knowledge in a hierarchical fashion, as was proposed by Charniak with simple vs. complex events and by Schank with MOPs, is applicable to the problems which we have encountered with frame selection. The result is an approach to frame selection which uses rules similar to those I presented in chapter 3.

The MOPTRANS parser uses MOP-like structures, which are language-independent structures used to represent the story as it is parsed. These frames are arranged hierarchically, according to their level of specificity, and thus allowing for shared knowledge between frames in the system. The hierarchy also provides information that

```
Script representation:

    $DOCTOR:
      HAVE-MEDICAL-PROBLEM + MAKE-APPT. + GO +
      DOCTOR-WAITING-ROOM + TREATMENT + PAY

    $LAWYER:
      HAVE-LEGAL-PROBLEM + MAKE-APPT. + GO +
      LAWYER-WAITING-ROOM + LEGAL-CONSULTATION + PAY

    $CAR-WASH:
      HAVE-DIRTY-CAR + GO + WAIT-IN-LINE +
      GET-CAR-WASHED + PAY

MOP representation:

                                M-GET-SERVICE
                               /             \
    M-PROFESSIONAL-OFFICE-VISIT               \
              /   \                            \
       M-DOCTOR    M-LAWYER                     M-CAR-WASH


    M-GET-SERVICE
      NEED-SERVICE + GO + WAIT + GET-SERVICE + PAY

    M-PROFESSIONAL-OFFICE-VISIT
      HAVE-PROBLEM + MAKE-APPT. + (GO) + WAITING-ROOM +
      (GET-SERVICE) + (PAY)

    M-CAR-WASH
      HAVE-DIRTY-CAR + (GO) + WAIT-IN-LINE +
      GET-CAR-WASHED + (PAY)

    M-DOCTOR
      HAVE-MEDICAL-PROBLEM + (MAKE-APPT) + (GO) + DOCTOR-WAITING-ROOM +
      TREATMENT + (PAY)

    M-LAWYER
      HAVE-LEGAL-PROBLEM + (MAKE-APPT) + (GO) + LAWYER-WAITING-ROOM +
      LEGAL-CONSULTATION + (PAY)
```

Figure 5-6: Script vs. MOP Representation of Various Events

can be used in the frame selection process. Instead of treating frame selection as a word disambiguation problem, as it was treated in request-based parsers, general frame selection rules are used in MOPTRANS, in conjunction with the hierarchical memory. The dictionary definition of a word points to a general concept in the hierarchy, which is

general enough to include all of the word's possible meanings. Then, general concept refinement rules can operate on the hierarchy, to refine the meaning of the word to a more specific frame.

To make this more clear, let us return to the police investigation example. In the request-based disambiguation method, it would be necessary to list in the dictionary definition of "realizar diligencias" all the possible frames to which the phrase could refer. In addition, a very large set of requests would be needed, to determine which frame was appropriate for a given context. However, in the MOPTRANS parser, the dictionary definition of "realizar diligencias" simply includes a pointer to a very general concept, called ACTION. In other words, this definition states that "realizar diligencias" refers to an action. Under the node ACTION in MOPTRANS' hierarchy are all of the frames in the system which represent actions. Concept refinement rules guide the selection of more specific frames, depending on the way in which the conceptual representation is built during the parse of the story. Figure 5-7 illustrates the placement of the possible frames to which "diligencias" can refer in the hierarchy.

Since the hierarchy of concepts used to refine the meaning of "diligencias" is language-independent, it is not used only for the disambiguation of "diligencias." Other words also point into the hierarchy at the appropriate level, depending on the specificity of their meaning, as is shown in Figure 5-7. Thus, "shop" points to a more specific node than "diligencias," but the same concept refinement rules are responsible for determining if either word refers to the structure GROCERY-STORE (even though the two words are from different languages).

```
                      X          "diligencias"
                    /   \             |
                  X      ACTION <---+        "shop"
                 / \      / \                  |
"find"          X     X     ATRANS             |  "grocery shopping"
  |                  / \    / \                |    |
  |            MTRANS  X SHOP <----------------+    |
  |              / \      / \                       |
  +---------> FIND  X  GROCERY-STORE <--------------+
"investigation"/ \
      |           X
      |          / \
      +----------> POLICE-INVESTIGATION
```

Figure 5-7: Hierarchical Structure of MOPTRANS' Conceptual Knowledge

For the police investigation story, the MOPTRANS parser uses some of the conceptual structures in Figure 5-7. In this diagram, the branches of the tree represent IS-A links. All of the concepts in this IS-A hierarchy have case frames, specifying the prototypical fillers for various slots, such as ACTOR, OBJECT, etc. For example, the case frame for FIND indicates that its ACTOR should be a PERSON, its OBJECT should be a PHYSICAL OBJECT, and its RESULT should be a GET-CONTROL. The case frame for POLICE-INVESTIGATION indicates that its ACTOR should be an AUTHORITY, its OBJECT should be a CRIMINAL, and its RESULT is an ARREST.

In addition to the hierarchical information, event sequences, similar to MOPs, are needed to represent knowledge about which of these frames are likely to occur together,

and what the causal relationships between the frames are. The two event sequences needed for this example are the following:

```
GET            = FIND + GET-CONTROL
POLICE-CAPTURE = CRIME + POLICE-INVESTIGATION +  ARREST
```

Note that the structure GET is part of the named plan USE in (Schank and Abelson, 1977)).

Recall the line of reasoning that I suggested in chapter 2 that a human reader might follow in order to infer that "realizar diligencias" means POLICE-INVESTIGATION in this example. First, since the prepositional phrase "para capturar" (in order to capture) follows "realizar diligencias," a human reader knows that the action expressed by "realizar diligencias" somehow will lead to a capture, or that the capture is the goal of the "diligencias." Capturing something involves getting control of it, and we know that before we can get control of an object, we have to know where it is and we have to find it. This indicates that perhaps "realizar diligencias" refers to some sort of finding. But when police are trying to find something in order to get control of it, they usually do a formal type of search, or an investigation. Therefore, we know that in this case, the word "diligencias" refers to a police investigation.

With the hierarchical memory organization and stereotypical event sequence knowledge presented in Figure 5-7 very general rules can be used to perform the frame selection for "realizar diligencias" along these same lines. First, the word "capturar" refers to the concept GET-CONTROL. From the event sequence GET above, we know that GET-CONTROL is often preceded by the event FIND. Since the story says that some action, "diligencias," precedes the GET-CONTROL, we can infer that the action is probably a FIND. This suggests the following general inference rule: If a scene of a script is mentioned in a story, then other scenes of the same script can be expected to be mentioned. Then, if an abstraction of another scene of the script is mentioned, we can infer that the abstraction actually is the other scene. In more concrete terms, in this example GET-CONTROL is a scene of the script GET. Another scene of GET is the scene FIND. "Realizar diligencias" refers to an abstraction of the concept FIND, namely ACTION. Since GET-CONTROL was mentioned, indicating that other scenes of the script GET are likely to be encountered, we can infer that the ACTION is actually a FIND, since ACTION is an abstraction of FIND.

Put more precisely, the inference that "diligencias" probably means FIND is performed by the following rules:

SCRIPT ACTIVATION RULE: If an action which is part of a stereotypical event sequence is activated, then activate the stereotypical event sequence, and expect to find the other actions in that sequence.

EXPECTED EVENT SPECIALIZATION RULE: If a word refers to an action which is an abstraction of an expected action, and the slot-fillers of the action meet the prototypes of the slot-fillers of the more specific action, then change the representation of the word to the more specific expected action.

Next, consider how we can infer that the FIND is a POLICE-INVESTIGATION. First, in the story the ACTOR of the FIND is the POLICE. One piece of knowledge that we have about POLICE is that often they are the ACTORs of POLICE-

INVESTIGATIONs, since that is part of their job. Then, since the IS-A hierarchy tells us that POLICE-INVESTIGATION is a refinement of the concept FIND, we can infer that in this story, the FIND is most likely a POLICE-INVESTIGATION.

This suggests the following inference rule:

> SLOT-FILLER SPECIALIZATION RULE: If a slot of concept A is filled by concept B, and B is the prototypical filler for that slot of concept C, and concept C IS-A concept A, then change the representation of concept A to concept C.

In this case, concept A is FIND, and concept B is the POLICE. The POLICE are the prototypical ACTORs of concept C, a POLICE-INVESTIGATION. Since FIND is above POLICE-INVESTIGATION in the IS-A hierarchy, then we can conclude that FIND in this case refers to POLICE-INVESTIGATION.

MOPTRANS uses these three general inference rules, the Script Activation Rule, the Expected Event Specialization Rule, and the Slot-filler Specialization Rule; to perform frame selection for "realizar diligencias" in the example above. These rules require the organization of knowledge structures in a hierarchical fashion, so that they can use this hierarchy to guide the refinement of concepts. They also require the existence of MOP-like structures in memory, to provide expectations as to what actions are likely to occur together in stories.

Given these rules, the disambiguation of "realizar diligencias" in the original police investigation example proceeds as follows: first, a general representation is built for "realizar diligencias"; simply, the concept ACTION. Then, the ACTOR of ACTION is filled in by an appropriate slot-filling rule (which will be discussed in the next chapter), which looks to the left of "realizar diligencias" for its ACTOR. This causes the concept AUTHORITY (the representation of "policia") to be filled in as the ACTOR of the ACTION. Next, the concept GET-CONTROL is built from the word "capturar." This also causes the event sequence GET to be activated, because of the Script Activation Rule. This, in turn, causes the concept ACTION to be changed to the concept FIND, due to the Expected Event Specialization Rule. Now, since the ACTOR slot of FIND is filled by AUTHORITY, and since the prototype of the ACTOR slot of POLICE-INVESTIGATION is AUTHORITY, the concept FIND is changed to be POLICE-INVESTIGATION because of the Slot-filler Specialization Rule.

Unlike the syntactic transfer rules in chapter 2 or the requests in chapter 4, these same frame selection rules apply to many rewordings of the police investigation story. This is because the rules are expressed *in purely conceptual terms*. Other parsing rules (which will be discussed in the next chapter) are responsible for filling in the slots in the representation of "realizar diligencias." This was not true with request-based rules, because requests were responsible for both filling in the slots of the representation *and* selecting the appropriate frame for "realizar diligencias." Thus, the requests were more example-specific.

To see that the general concept refinement rules apply to rewordings of this story, consider the following story:

> Spanish: INTENSAS DILIGENCIAS POR PARTE DE LA POLICIA resultaron en la captura de un reo.
>
> English: AN INTENSE POLICE INVESTIGATION resulted in the arrest of a criminal.

Here, MOPTRANS goes through the same procedure to select the frame POLICE-INVESTIGATION. First, AUTHORITY is assigned to be the ACTOR of the ACTION referred to by "diligencias." This information is supplied by the attachment of the prepositional phrase "por parte de la policia" to "diligencias" (the way in which this attachment proceeds will be discussed in the next chapter). Then, the verb "resultaron" provides the information that the ACTION done by the POLICE is IN-SERVICE-OF a GET-CONTROL ("captura"). Again, the activation of the concept GET-CONTROL also causes activation of the event sequence GET, because of the Script Activation Rule. Next, the concept ACTION is refined to be FIND, just as before, because of the Expected Event Specialization Rule. Now, since the ACTOR slot of FIND is filled by AUTHORITY, and since the prototype of the ACTOR slot of POLICE-INVESTIGATION is AUTHORITY, the concept FIND is changed to be POLICE-INVESTIGATION because of the Slot-filler Specialization Rule.

Let us return to another example from chapter 2, which was problematic for syntax-based systems. These examples involved the translation of the word "ganar":

Spanish: Yo GANE mil dolares en la noche del ano nuevo en el casino.

English: I WON one thousand dollars on New Year's eve at the casino.

Spanish: En el casino los talladores GANARON mil dolares en la noche del ano nuevo cada uno.

English: At the casino the dealers each EARNED one thousand dollars on New Year's eve.

Spanish: Los talladores que trabajaron en el casino en la noche del ano nuevo GANARON mil dolares cada uno.

English: The dealers who worked on New Year's eve at the casino each EARNED one thousand dollars.

The MOPTRANS parser can correctly translate "ganar" in these examples, using general concept refinement rules. The three examples share common situations, which can be captured in the following structures:

```
          ATRANS
          /    \
       WIN  GET-PAID
```

```
EMPLOYMENT:  DO-JOB  +  GET-PAID (earn)
                                            WIN
BET:         PLACE-BET  +  PLAY-GAME  +     or
                                            LOSE
```

Given the event sequences EMPLOYMENT and BET, it is an easy matter for the MOPTRANS parser to formulate an expectation to find either a GET-PAID scene or a WIN scene, depending on which event structure is active. This expectation, in conjunction with the hierarchical knowledge above linking ATRANS with GET-PAID and WIN, is used to select the right frame for "ganar," which is defined as a type of ATRANS.

To facilitate the instantiation of the event sequences, a few more event sequence

instantiation rules are needed:

EVENT SEQUENCE LOCATION INSTANTIATION RULE: If a setting or location is mentioned which is associated with a particular event sequence, and a person who would be likely to take part in that event sequence is mentioned, then instantiate the event sequence.

ACTOR LOCATION INSTANTIATION RULE: If a person is in a location in which he typically engages in a particular event sequence, then instantiate that sequence.

INSTANTIATION PRECEDENCE RULE: If both of the above rules apply in a story, use only the Actor Location Instantiation Rule.

These six general rules allow the MOPTRANS parser to disambiguate the word "ganar" for the 3 examples above. In the first example, "I won a thousand dollars on New Year's Eve at the casino", the Event Sequence Location Instantiation Rule applies, since a casino is a place where betting occurs. BET is instantiated, and then the Expected Event Specialization Rule applies, since BET expects to find WIN in the story. In the second and third examples, the Actor Location Instantiation Rule applies, since it has precedence over the Event Sequence Location Instantiation Rule, and EMPLOYMENT is instantiated. Since EMPLOYMENT expects GET-PAID as a scene, the Expected Event Specialization Rule applies, and GET-PAID is instantiated.

The knowledge structures and concept refinement rules I have outlined here are by no means enough to translate "realizar diligencias" or "ganar" correctly in all contexts, but they do allow for the six concept refinement rules above to choose the correct translations in these examples. The problems involved with formulating a set of rules which would work in all cases are quite difficult. However, a hierarchical memory structure does provide a good framework for writing rules such as the above ones which can accurately distinguish between a limited number of meanings of a word within a limited domain. This frame selection ability can be accomplished without the proliferation of rules which were encountered using syntax-based methods or conceptual methods with lexically-based disambiguation rules.

To emphasize that this sort of concept refinement process can be used often in natural language processing, let us examine one more example in which this process takes place. It involves the word "seized":

Iranian students seized control of the American Embassy in Tehran.

A gunman seized control of a Boeing 727 and diverted it to Cuba.

A gunman seized three people as hostages and demanded a $5 million ransom.

"Seized" is a sufficiently vague word in the domain of terrorism and crime to require several word senses in the request-based method of disambiguation. In the examples above, "seized" refers to the frames TAKE-OVER-BUILDING, HIJACK, and TAKE-HOSTAGES. Thus, a request-based system would require three separate requests for these sentence, looking to the right of the verb for a BUILDING, a VEHICLE, or a PERSON. However, in MOPTRANS, "seized" is defined as having only one sense, meaning GET-CONTROL. All of the more specific frames to which "seized" could refer are under GET-CONTROL in the hierarchy. Thus, the slot-fillings of the ACTOR and OBJECT slots of GET-CONTROL cause the appropriate frame to be selected by the concept refinement rules. If the OBJECT of the GET-CONTROL is a BUILDING, then

the frame TAKE-OVER-BUILDING is chosen, because the slot-filler BUILDING matches the prototype for the OBJECT slot of TAKE-OVER-BUILDING, which IS-A GET-CONTROL[8]. Similarly, if the OBJECT is filled with a VEHICLE, then the system would choose the frame HIJACK, because the prototypical OBJECT for a HIJACK is a VEHICLE. The same is true of "hostages," which matches the prototype for the OBJECT of TAKE-HOSTAGES.

The economy of concept refinement rules over requests can be illustrated further with the following sentences:

> The seizing of the American Embassy by Iranian students took place yesterday.
>
> Passengers on a Boeing 727 seized by a gunman and diverted to Cuba were freed after the gunman was overpowered by the pilot.
>
> Police arrested a gunman who seized three people as hostages and demanded a $5 million ransom.

Since the operation of the concept refinement rules in MOPTRANS do not depend on the syntactic construction of a sentence, the same concept refinement process used in the first three sentences would handle these three sentences. However, this is not the case with requests. For the first example above, additional requests would be required to look to the right of the preposition "of" for a BUILDING, a VEHICLE, or a PERSON. In the other two examples, the requests determining whether or not "seized" is past active or unmarked passive would have to be duplicated for each case, looking for a BUILDING, a VEHICLE, or a PERSON. Thus, a great number of additional requests would be required for these examples.

5.4 Concept Refinement Rules in MOPTRANS

5.4.1 More About the Hierarchy

The hierarchical organization of knowledge which I have discussed is encoded in the MOPTRANS parser in terms of IS-A pointers, which point from a structure to more abstract structures. Thus, part of the definition of a conceptual structure in the MOPTRANS parser is a pointer to its ancestor in the hierarchy. For example, the structure SHOOT points to a more abstract structure, called HARM.

A structure can have IS-A links to more than one abstract structure. Thus, the data structure in which conceptual structures are stored is not really a hierarchy, but rather a directed, acyclical graph. For example, the structure ESCAPE is a type of GET-CONTROL, where the ACTOR of the ESCAPE is taking control of himself from the person who had control of him. Thus, ESCAPE has an IS-A link to GET-CONTROL.

---

[8]It could be that there are other frames in the system that are GET-CONTROL's whose OBJECT is a BUILDING. For instance, if the system contained a structure like FORECLOSURE, this action would have a prototype of BUILDING for its OBJECT slot, too. However, the additional information that the ACTOR of this action is "Iranian students" would still cause the frame TAKE-OVER-BUILDING to be selected, because the ACTOR slot-filler would violate the prototype for the ACTOR of a FORECLOSURE.

However, ESCAPE also has an IS-A link to PTRANS, since ESCAPE also involves transfer location of oneself away from one's captor.

Properties of structures are inherited down IS-A links. For example, the structure ATRANS has certain slots: ACTOR, OBJECT, FROM, and RECIPIENT. These slots are expected to be filled with a PERSON, a PHYSICAL-OBJECT, a PERSON, and a PERSON, respectively. This is then true of all actions with ATRANS as an ancestor. Every such action has (at least) the slots ACTOR, OBJECT, FROM and RECIPIENT. The prototypical fillers of these slots are at least as specific as the prototypical fillers for the action PTRANS. Thus, the ACTOR of an action whose ancestor is PTRANS is at least as specific as PERSON, and may be some subset of the class PERSON (e.g., the ACTOR of an ARREST, which has ATRANS as an ancestor, is a POLICE, which IS-A PERSON.)

One of the possible properties of a conceptual structure may be a pointer to an event sequence which the structure is a part of. For example, the structure ARREST points to the event sequence POLICE-CAPTURE, which also contains the events CRIME and POLICE-INVESTIGATION. Event sequences also point to ancestors, if a more abstract version of the event sequence exists. For example, POLICE-INVESTIGATION points to the structure GET, which consists of only two events, FIND and GET-CONTROL. The events which make up one event sequence which is an abstraction of another event sequence must be abstractions of the events in the more specific event sequence. For instance, in ARREST, FIND is an abstraction of POLICE-INVESTIGATION, and GET-CONTROL is an abstraction of ARREST. POLICE-CAPTURE contains an additional event, CRIME, for which there is no corresponding event in GET.

5.4.2 How Concept Refinement Works

The six concept refinement rules which I discussed above, which operate on this hierarchy of knowledge, are implemented as demons in the MOPTRANS parser. Some of these demons inspect new conceptualizations whenever one is built. If a conceptualization is built which satisfies the conditions of one of these rules, then the demon instantiates the appropriate event sequence. For example, when the concept GET-CONTROL is built in the police investigation examples, the demon corresponding to the Script Activation Rule builds an instantiation of the event sequence GET.

These demons must use the IS-A links provided in the hierarchy during their checks. For example, the concept GET-CONTROL points to the event sequence GET. But if a story referred to a more specific concept, such as STEAL, the sequence GET should still be activated, since before stealing something, one must find it. Thus, newly built conceptualizations must be examined to see if they point to an event sequence, or if any concepts further up in the hierarchy point to event sequences.

The two other inference rules from chapter 4 were concept refinement rules, specifying conditions under which the parser could change the representation of an object or an event to a more specific representation:

EXPECTED EVENT SPECIALIZATION RULE: If a word refers to an action which is an abstraction of an expected action, and the slot-fillers of the action meet the prototypes of the slot-fillers of the more specific action, then change the representation of the word to the more specific expected action.

SLOT-FILLER SPECIALIZATION RULE: If a slot of concept A is filled by concept B, and B is the prototypical filler for that slot of concept C, and concept C IS-A concept A, then change the representation of concept A to concept C.

The implementation of the Expected Event Specialization Rule is in the form of two demons. The first, which is just like the demons for the event sequence activation rules above, examines newly instantiated conceptualizations to see if they are more general versions of expected actions. The second demon is activated when a new event sequence is built, to see if already-built conceptualizations are possible members of the event sequence. This second demon performs the refinement of the ACTION representing "diligencias" to FIND, since "diligencias" appears in the police investigation example before "capturar," which builds GET-CONTROL and causes the instantiation of the event sequence GET.

The implementation of the Slot-filler Specialization Rule is also as a demon, which inspects conceptualizations whenever a slot-filling is performed by the parser. However, recognizing that the conditions of this demon have been met is somewhat trickier. This is because the demon must know whenever the new filler of a slot meets the prototype for that slot of ANY of the frames in the system which have IS-A pointers to the current frame. For example, when the concept FIND in the police investigation story is built, its ACTOR is assigned to be the POLICE. The Slot-filler Specialization demon must realize that POLICE is the prototypical ACTOR of the more specific concept, POLICE-INVESTIGATION. To do this, it seems that this demon must inspect the case frames of every single concept which is a FIND. In general, the inspection of the case frames of all concepts which are more specific versions of a given concept could be quite costly.

To make the search that this demon must perform more efficient, conceptualizations are indexed in the MOPTRANS parser according to the slots in their case frames, as well as the expected prototypes for the fillers of these frames. Thus, one way in which POLICE-INVESTIGATION is indexed is by the slot ACTOR, and the expected filler POLICE. Then, when the concept FIND is assigned to have the ACTOR POLICE, the demon is able to find the concept POLICE-INVESTIGATION through the indices ACTOR and POLICE.

Actually, the search process is not that simple, due to three complications. First, it may be that this indexing process will find frames which are not more specific versions of the current frame. For example, another action whose ACTOR is typically the POLICE is GIVE-TICKET. If this frame were in the system, the indexing mechanism would find it. Thus, one additional check that the demon must make is that the frame has IS-A links to the current frame.

A second problem is that the slot-filler concept may not point directly to the desired frame. Instead, a more general concept may point to this frame. For example, if the ACTORs in the police investigation story were the FBI, the index *FBI* would not find the frame POLICE-INVESTIGATION. This is because the prototypical ACTOR of POLICE-INVESTIGATION is not that specific. Thus, in addition to using the slot-filler concept as an index, concepts further up the IS-A hierarchy must be used, also.

Finally, a third problem is that more than one frame might be found by the indexing process. If this happens, then the demon may or may not be able to refine the current frame. Two situations will illustrate when a frame should and should not be chosen, given more than one frame retrieved by the indexing process. In the police investigation example, which I will call situation 1, when the parser initially assigns POLICE to be the

ACTORs of the ACTION, many frames are found by the indices ACTOR and POLICE, all of which are more specific versions of the current frame, ACTION. Some of these frames would be GIVE-TICKET, POLICE-INVESTIGATION, ARREST, etc. In this case, none should be chosen by the demon, since there is not enough information to determine which is the right frame.

However, consider situation 2, an example discussed in (Schank, Birnbaum, and Mey, 1983):

John got a TV at Macy's.

Given the slot-filler "Macy's" as the LOCATION of the ATRANS representing "got," we can infer that John *bought* the TV. Thus, in this situation the parser should refine ATRANS to the more specific frame, BUY. However, it may be that more specific frames exist in the parser which could possibly apply. For example, the frames CREDIT-CARD-BUY and CASH-BUY might exist. If this were so, these frames would also be found by the indexing process.

To allow for concept refinement to occur in situation 2 but not in situation 1, the Slot-filler Specialization demon chooses a frame from a group of frames found through indexing only if a path can be found to that frame from all of the other frames found, via IS-A links. We can see that this selection heuristic works by examining the graphic representations in Figure 5-8 of the two situations above. In situation 1, the only node which dominates all of the candidate frames is the current frame, ACTION. Thus, no concept refinement should take place in this case. However, in situation 2, BUY dominates both CREDIT-CARD-BUY and CASH-BUY in the IS-A hierarchy. Thus, the Slot-filler Specialization demon refines ATRANS in this situation to the concept BUY.

```
        ACTION                              ATRANS
         / \                                /    \
        X   X                              X      X
       / \ / \                            / \    / \
      X     POLICE-INVESTIGATION              BUY
     / \                                      / \
    X  GIVE-TICKET                           X  CASH-BUY
   / \                                      / \
    ARREST                                   CREDIT-CARD-BUY
```

Figure 5-8: Hierarchical Arrangement of the Frames in Situations 1 and 2

5.5 Vagueness vs. Genuine Ambiguity

There are two types of words for which frame selection is an issue: vague or general words, and what I will call "genuinely" ambiguous words. "Realizar diligencias" is an example of a vague word or phrase. By this I mean that the different possible meanings of this phrase all have something in common semantically. Of course, this is trivially true of all ambiguous words; if nothing else, all of the meanings of an ambiguous word refer to a concept. However, in the case of vague words, all the possible meanings of a vague word are colorations of a common concept, and also most possible colorations of that

concept can be referred to by the vague word. So, in the case of "realizar diligencias," all of this phrase's possible meanings are "diligent actions". What's more, most diligent actions can be expressed in some way in Spanish by using the phrase "realizar diligencias." Thus, by this definition, the phrase is vague.

With "genuinely" ambiguous words, on the other hand, the different possible meanings of a word do not necessarily share a common abstract meaning, or if they do, not all the possible colorations of that abstract meaning can be expressed using the ambiguous word. For example, the verb "to cry" can refer to shouting, or to the crying of tears. One might think that "cry" is a vague word, since both actions are types of MTRANS's (in some sense). However, not every type of MTRANS can be expressed using the word "cry." For example, whispering, a type of MTRANS, cannot be expressed using the word "cry." Thus, "cry" is a genuinely ambiguous word.

Frame selection for these two types of ambiguous words is implemented in the MOPTRANS parser in slightly different ways. In the case of vague words, the dictionary definition of the word consists simply of a pointer to a structure in the conceptual hierarchy which reflects the level of vagueness of the word, along with any additional stipulations on meaning which that word conveys. For example, the phrase "realizar diligencias" points to the structure *DO*, indicating that it must refer to an intentional action. This structure is relatively high up in MOPTRANS's conceptual hierarchy, reflecting the extreme vagueness of this phrase. An additional feature that the action is diligent would also be stipulated, in the form of some additional slot-filling information. Then, resolution of vagueness of the word is performed by the demons described above.

For genuinely ambiguous words, a pointer to a structure will not suffice. For vague words, this is enough, due to the fact that a vague word can refer to any descendant of the node pointed to by the word, and that any node could conceivably be reached by the execution of the demons. However, genuinely ambiguous words cannot refer to every possible coloration of a concept. For this type of word, the dictionary definition consists of several pointers into the hierarchy, corresponding to each of the words distinct meanings. These pointers function as IS-A links within the hierarchy. Thus, the dictionary definition acts as a "dummy" node within the IS-A hierarchy, with IS-A links added from every possible meaning of the ambiguous word to the dummy node. When a genuinely ambiguous word is read by the MOPTRANS parser, its initial representation is simply a pointer to its dictionary definition. Then, the same frame selection process that is used for vague words can be used, since the concept refinement demons will eventually refine from the dummy node to a real concept in the hierarchy.

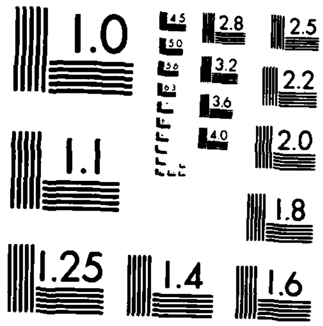
To make this more clear, consider the verb "to fix." It has (at least) two distinct meanings, corresponding to the following two uses:

John fixed the washing machine.
John fixed the horse race.

To distinguish between these meanings, the dictionary definition of "fix" would have two pointers into the conceptual hierarchy, one to the node REPAIR, and the other to the node RIG. REPAIR would expect a PHYSICAL OBJECT as its semantic OBJECT, while RIG would expect some sort of ACTION as its OBJECT. In the two examples above, a dummy structure would first be built to represent "fix." Then, syntactic rules which will be discussed in the next chapter would assign either "washing machine" or "horse race" as the OBJECT of "fix." This would cause the Slot-filler Specialization demon to choose either REPAIR or RIG, because the slot-filler would meet the prototype for only one of the frames REPAIR and RIG. Thus, the correct frame would be selected

AD-A150 505 THE ORGANIZATION OF KNOWLEDGE IN A MULTI-LINGUAL INTEGRATED PARSER(U) YALE UNIV NEW HAVEN CT DEPT OF COMPUTER SCIENCE S L LYTINEN NOV 84 YALEU/CSD/RR-340 2/3

UNCLASSIFIED N00014-82-K-0149 F/G 5/7 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

on the basis of the semantic properties of the direct object of "fix."

To emphasize the advantages of this sort of frame selection technique over the request-based method which I discussed in chapter 3, notice that this same hierarchical information could easily disambiguate nominalized forms of "fix," as in the following examples:

> The fixing of the horse race was done by the mob.
> The fixing of the washing machine took several hours.

During the parse of the two examples, the semantic OBJECT of "fix" is assigned by a syntactic rule which recognizes the pattern of a present participle followed by the preposition "of," followed by the present participle's semantic OBJECT. Once this semantic slot-filling is performed, the same hierarchical information that caused the disambiguation to occur in the earlier examples resolve the ambiguity in this situation.

5.6 Using Concept Refinement Demons for Prepositions and Adjectives

Two classes of words which are often semantically ambiguous are prepositions and adjectives. The MOPTRANS parser disambiguates both of these classes of words using the concept refinement techniques described above.

Prepositions are often vague or ambiguous, referring to many possible semantic relationships. For example, the word "for" was shown by (Hemphill, 1975) to have 20 different meanings, referring to semantic relations such as IN-PLACE-OF, DURATION, and RECIPROCAL-CAUSALITY.

To handle the vagueness of prepositions, they are defined as any other vague or ambiguous word in the MOPTRANS parser is defined. The dictionary definition consists of one or more pointers which point to a semantic relation, defined in the conceptual memory of the parser, which are at the appropriate level of vagueness or generality. These semantic relations are just like other concepts in the MOPTRANS parser, with slots which can be filled in and prototypes for what semantic concepts can fill those slots. They are also arranged hierarchically, just as other concepts are arranged. As the parser fills in the representation and fills these slots, the concept refinement demons operate on the structure, just as they would with any other structure.

Semantic relations are always defined to have (at least) two slots, called S1 and S2. These slots correspond to the conceptualization in which the relation appears, and the slot-filler which fills this slot in the conceptualization. For example, in "John gave the book to Mary," the preposition "to" refers to the relation RECIPIENT. The S1 slot of RECIPIENT is filled with the concept ATRANS, built by "gave"; and the S2 slot of RECIPIENT is filled by (HUMAN GENDER FEMALE NAME MARY).

Let us consider an example of an ambiguous preposition which the MOPTRANS parser handles. The preposition "in" can refer to many relations, including the following:

```
The shooting in the town ...          LOCATION
The soldier shot in the arm ...       HURT-PART
He was killed in a raid ...           DURING
The first killing in 3 years ...      AFTER
```

"In," in all of these examples, specifies a semantic relation between the object of the

preposition and the noun group or verb which "in" is attached to. Thus, the LOCATION of the shooting was a town in the first example; the HURT-PART of the soldier in the shooting was his arm in the second example; the killing took place DURING a raid in the third example; and the killing in the last example took place AFTER another (inferred) killing, by 3 years.

To handle this ambiguity, the dictionary definition of "in" has pointers to all of the relations mentioned above. These pointers all function as IS-A links, so that the concept refinement rules can choose which relation "in" refers to, depending on the semantic context. This choice is made when the two slots of the dummy node built for "in" are filled in. Thus, if slot S1 of the dummy node is filled with the action SHOOT and slot S2 is filled with TOWN, as in the first example, the semantic refinement process chooses the relation LOCATION, since that relation is the relation whose prototypical slot-fillers best match the actual slot-fillers. Similarly, filling slot S2 with a BODYPART in the second example causes the relation HURT-PART to be chosen, since its S2 prototype is a BODYPART.

Adjectives are handled in much the same way in the MOPTRANS parser. Often, adjectives provide a conceptualization which will fill a slot, and a vague notion of what slot should be filled by this conceptualization. For instance, consider the following uses of the word "Chinese," functioning as an adjective:

```
a Chinese man                 (MAN NATIONALITY CHINA)
                              (MAN ANCESTRY CHINA)
the Chinese government        (GOVERNMENT CONTROL-OVER (NATION NAME CHINA))
Chinese pottery               (POTTERY ORIGINATION CHINA)
```

All of these uses of the word "Chinese" indicate that some property of the noun which "Chinese" modifies has to do with the country China. However, the particular property varies in each use of the word, from NATIONALITY or ANCESTRY to ORIGINATION and even CONTROL-OVER.

To handle the ambiguities of adjectives like "Chinese," these adjectives are defined in a similar way to prepositions. The dictionary definition of "Chinese" has pointers to all of the possible relations to which it could refer. The definition also specifies that the S2 slot will be filled with the conceptualization (NATION NAME CHINA), signifying that some property of the noun which the adjective modifies will be filled with this conceptualization. Depending on the conceptualization which fills the S1 slot, the concept refinement process chooses one of the possible meanings of "Chinese." For instance, if a PERSON fills the S1 slot, the NATIONALITY relation is chosen. However, if a PHYSICAL-OBJECT, like "pottery," fills the S1 slot, then ORIGINATION is chosen as the meaning of "Chinese."

5.7 Comparison to Other Work

5.7.1 Expectations from Other Frames

It is worth noting some similarities between the MOPTRANS parser's frame selection techniques and some other work done in frame selection. The MOPTRANS approach is similar in some ways to the approach used in the Integrated Partial Parser (IPP) (Lebowitz, 1980), which parsed short newspaper articles about terrorism; and in the GUS system (Bobrow, 1977), a system which conversed about airplane trips. In these systems, frames already selected were responsible for predicting other frames that were likely to appear in a text. These predictions helped to disambiguate words which could refer to many different frames. For example, in IPP the word "held" could refer to many different scripts: $TAKE-HOSTAGES, $TAKE-OVER (a building), and $KIDNAP. However, expectations from already active structures often determined which of these scripts "held" referred to. Thus, if the structure $HIJACK, another frame in IPP, was already active, then "held" was assumed to mean $TAKE-HOSTAGES, since hijackings often involve the taking of hostages.

This approach to frame selection is similar to the Expected Event Specialization Rule used in MOPTRANS. However, rules corresponding to the other concept refinement rules in MOPTRANS were not present in IPP and in GUS. Thus, frame selection in these system was incomplete, in that it was difficult to select an initial frame. If no frames were active at the beginning of a story, then no predictions could be made as to what other frames would occur in the story. Thus, if a structure like $HIJACK was not already active when "held" was encountered in IPP, then more traditional lexically-based requests would have to be used to choose a frame.

To avoid the problem of selecting an initial frame, the GUS system only dealt with texts having to do with airplane trips. Thus, the trip specification frame was always active at the beginning of the story. This frame could then be used to predict other frames that might appear in the text. The IPP parser also relied in part on a restricted domain to deal with the problem of selecting an initial frame. Many words in English which are vague in general are unambiguous in the domain of terrorism, and thus were unambiguous in IPP. For instance, the word "divert" in IPP referred to only one frame, namely $HIJACK. Lebowitz suggested that the restriction on meanings of ambiguous words by domain could actually be used as an approach to disambiguation, even when working with less restricted domains. If the parser could identify what domain a story belonged to, then it could use the domain to restrict the meanings of words in the story. However, he did not suggest how this might be done.

5.7.2 Frame Selection by Process of Elimination

A different approach to frame selection was presented in (Hirst, 1983). Hirst used what he called *Polaroid Words* to disambiguate semantically ambiguous words, provided all the possible uses of a word were of the same syntactic class. In his approach, the dictionary entry of an ambiguous word contained a list of all of its different possible meanings. At parse time, a Polaroid Word was built for each ambiguous word in a sentence. Each Polaroid Word was responsible for eliminating all but one of its word's possible senses, by means of testing each sense's compatibility with the surrounding context. To enable this, Polaroid Words communicated with each other in limited ways. When one possible meaning of a word was eliminated, the Polaroid Word responsible for

the word communicated this to other Polaroid Words, which in turn used this information to try to eliminate possible meanings of their ambiguous words. Thus, possibilities were gradually eliminated, until the disambiguation process was complete.

An example which Hirst presented was the sentence "The slug operated the vending machine," in which both "slug" and "operated" were ambiguous words. Their dictionary definitions were the following[1]:

```
[slug (noun):                        [operate (verb):
    gastropod-without-shell              [cause-to-function
    bullet                                  agent       SUBJ
    metal-stamping                          patient     SUBJ, OBJ
    shot-of-liquor]                         instrument  SUBJ, with]
                                         [perform-surgery
                                            agent       SUBJ
                                            patient     upon, on
                                            instrument  with]
```

The dictionary definition of "operate," in addition to providing a list of its possible meanings, also provided information as to where the semantic cases of the frames that it could refer to could be found. Thus, if "operate" meant PERFORM-SURGERY, then its subject would fill the AGENT case, its PATIENT would follow the preposition "upon" or "on," etc.

Hirst's parser used pseudo-prepositions, SUBJ and OBJ, inserted before the subject and object of the sentence. These pseudo-prepositions were treated as regular words, and were defined in the dictionary according to the semantic cases that they could mark. Since they could mark more than one case, they too were ambiguous. Here are their dictionary definitions:

```
[SUBJ (prep):                          [OBJ (prep):
    agent      animate                     patient     thing
    instrument physobj                     transferee  physobj]
    patient    physobj]
```

The disambiguation process worked as follows: first, "operated" provided the information to SUBJ that if SUBJ marked the AGENT case, the noun phrase the followed would have to be HANIM (higher animate). Since "slug" could not refer to a HANIM, SUBJ used this information to conclude that it did not refer to AGENT, leaving INSTRUMENT and PATIENT as possibilities. Next, since the definition of "operate" specified that SUBJ would flag the AGENT case if "operate" meant PERFORM-SURGERY, this meaning of "operated" could be eliminated, since SUBJ had already eliminated AGENT as a possible meaning. Thus, "operated" meant CAUSE-TO-FUNCTION.

Once "operated" was disambiguated, OBJ knew that it must mark the case PATIENT, due to case information from the CAUSE-TO-FUNCTION definition of "operate." Since cases could only be marked once in a sentence, this provided SUBJ with

---

[1]The dictionary definitions shown here are slightly simplified, with some portions that are irrelevant to this example left out.

enough information to conclude that it must refer to INSTRUMENT. Finally, "slug" was disambiguated by a different mechanism, called *marker-passing*, which found a path between the METAL-STAMPING sense of "slug" and "vending machine."

Like MOPTRANS, Hirst's approach to word disambiguation avoids the problems of mixing disambiguation knowledge with syntactic knowledge. Thus, Hirst did not need special-purpose rules which only applied to a particular ambiguous word, as is the case with requests. In addition, Polaroid Words appear to be a good approach to dealing with sentences containing more than one ambiguous word. However, Hirst did not offer a solution to the problem of disambiguating vague words. In Hirst's approach, if a word referred to a frame, the frame had to be listed in the dictionary entry of the word. Thus, vague words like "diligencias" would be difficult to disambiguate using Hirst's approach.

5.7.3 Frame Selection by Discrimination

MOPTRANS' frame selection approach is also similar to that used in the FRUMP system (DeJong, 1979), but with some advantages over FRUMP's system. FRUMP produced summaries of newspaper articles from many domains. Thus, the frame selection problem was very real in FRUMP. To handle this problem, DeJong used discrimination nets called sketchy script initiator discrimination trees (SSIDTs). One SSIDT existed for each Conceptual Dependency primitive. An SSIDT, when given a Conceptual Dependency representation, selected a frame, or "sketchy script," on the basis of the roles and role fillers contained in the CD representation. Thus, a text was first decomposed into its CD representation, then parsing rules would fill in various roles in the representation, and finally an SSIDT selected a sketchy script on the basis of what roles were filled in, and how they were filled.

SSIDT's selected the sketchy script $EARTHQUAKE for the word "trembled," as in "The ground trembled." First, the word "trembled" was represented by PTRANS, the CD primitive for physical motion. In addition, "trembled" provided the information that the motion was cyclical in manner. Then, parsing rules assigned "ground" to be the OBJECT of this PTRANS. Finally, the SSIDT consisted of the following:

```
                    PTRANS
                      |
                node 2 (OBJECT)
               /      |      \
          GROUND   VEHICLE   HUMAN
            |
      node 3 (ACTOR)
          /   \
  EXPLOSIVE   GEOLOGICAL FORCE
                    |
              node 4 (MANNER)
                    |
                 CYCLICAL
                    |
               $EARTHQUAKE
```

Thus, the role-fillers of PTRANS, in this case the fact that the OBJECT was the ground and the MANNER of the motion was cyclical, guided the SSIDT to the sketchy script $EARTHQUAKE.

The SSIDT's which DeJong used are similar to the hierarchical organization of knowledge which is used in MOPTRANS. However, MOPTRANS' frame selection

method has several advantages over FRUMP's. First, in MOPTRANS' frame selection method, text does not need to be represented in terms of Conceptual Dependency primitives at the beginning of the frame selection process. Words can build any type of frame, from very general, as was the case with the phrase "realizar diligencias," to very specific. This was not true in FRUMP, as SSIDT's were indexed under the CD primitives. Thus, FRUMP had the restriction that a word's meaning must first be represented at the level of Conceptual Dependency primitives. This works well for words such as "trembled," which clearly refer to a particular primitive. However, in the case of very vague words, such as "diligencias," the initial representation in terms of CD would be problematic. It is not clear which primitive "diligencias" refers to. In fact, "diligencias" could conceivably refer to actions which would be reprsented by any of the CD primitives. Likewise, for very specific words which refer to specific frames, such as "kidnap," the restriction that the word first be represented in terms of CD is cumbersome. Instead of representing "kidnap" initially as an ATRANS, the system ought to be able to immediately find the frame KIDNAP without using an SSIDT as an index.

Second, although MOPTRANS' organization of frames in a hierarchy serves much the same function as the discrimination nets used by DeJong, the traversal of the hierarchy in the approach I have presented is less *ad hoc* than in FRUMP. The definitions of case frames themselves provide the discrimination rules for traversal of the net. In DeJong's system, arbitrary tests were used to determine what nodes in the discrimination net should be traversed. In the "diligencias" examples above, MOPTRANS was able to determine that FIND was actually a POLICE-INVESTIGATION because the case frame of POLICE-INVESTIGATION stated that the ACTOR of a POLICE-INVESTIGATION is the POLICE. This information, in conjunction with the slot-filler specialization rule, rather than an arbitrary discrimination rule, allowed the MOPTRANS parser to make the inference that the POLICE-INVESTIGATION frame was the most appropriate one for the context.

5.7.4 Taxonomic Lattices

On the surface, the frame selection process which I have described here is also similar in some respects to the Incremental Description Refinement process used in RUS (Bobrow and Webber, 1980). In this system, a taxonomic lattice (Woods, 1978) is used to refine the semantic interpretation of a sentence as it is being parsed. The refinement process is similar to the frame selection method I have described here in that it relies on the structure of a hierarchy to provide it with the information needed to discriminate to more specific concepts in the hierarchy. For example, the sentence "John ran the drill press" was parsed in this system using a taxonomic lattice containing nodes RUN-CLAUSE, PERSON-RUN-CLAUSE, RUN-MACHINE-CLAUSE. The parser refined its semantic interpretation of the sentence from RUN-CLAUSE to the more specific PERSON-RUN-CLAUSE and finally RUN-MACHINE-CLAUSE as more information was provided by the parse of the sentence.

However, there are many substantial differences between the RUS system and MOPTRANS. Although the refinement process itself and the structure of the hierarchies used in the two systems are similar, the content of the nodes in these hierarchies is completely different. First, the nodes in the taxonomic lattice in RUS are in no way independent of lexical items. Thus, the node RUN-MACHINE-CLAUSE would be distinct from OPERATE-MACHINE-CLAUSE, or nodes corresponding to other verbs which can refer to the operation of a machine. This is in contrast to the nodes in the hierarchy in

MOPTRANS, which are meant to be elements in a conceptual representational system. Second, since the nodes in RUS's taxonomic lattice are not meant to be conceptual representations, RUS contains no script-like knowledge about likely sequences of nodes. In contrast, the frame selection system which I have presented here also makes use of script-like sequences of events, which are meant to represent conceptual facts about the world. The information provided by this scriptal knowledge is an important part of the frame selection process in MOPTRANS.

5.8 Conclusion

In this chapter, I have presented six general inference rules which can be used to perform frame selection for sentences containing vague or general words. These rules draw on information from a hierarchically organized conceptual memory, which provides knowledge about abstractions of events and sequences of events.

This frame selection method is in sharp contrast to the lexically-based disambiguation methods which have been dominant in previous conceptual parsers, and it avoids the problems of rule explosion that are prevalent in these parsers. In the lexically-based request method, at least request is needed for each sense of an ambiguous word. Thus, using lexically-based requests to disambiguate very vague or general rules results in an explosion in the number of rules needed. On the other hand, the frame selection method used in the MOPTRANS parser does not suffer from the same rule explosion, because only general inference rules are used to perform disambiguation. Other knowledge necessary for this process is represented in a non-linguistic form, and thus does not need to be duplicated over and over again in the form of lexically-based rules, as was the case with requests.

Since most of the knowledge in the MOPTRANS system used for frame selection is represented in the hierarchically-organized conceptual knowledge base, rather than in language-specific rules, the MOPTRANS frame selection method has additional advantages. First, the frame selection knowledge used in the system is applicable to all of the natural languages that MOPTRANS parses. Since the same hierarchy of concepts is used in MOPTRANS no matter what the source or target language, this knowledge is available for disambiguating words in any language. This would not be true in a multi-lingual request-based parser, since conceptual knowledge in such a parser would be largely lexically-based, and therefore not easily shared across languages.

Second, the negative implications of learning in parsers using lexically-based rules do not apply to the organization of knowledge in the MOPTRANS system. Recall that in the Word Expert Parser, knowledge used to disambiguate "throw" seemed like it should be applicable to tasks other than parsing, such as a vision system watching someone throw an object. Thus, any knowledge learned in parsing should apply to vision processing, and vice versa. However, since this knowledge was stored in the lexicon in the Word Expert Parser, it was difficult to imagine how any knowledge learned for parsing could apply to other tasks. This is not the case in the MOPTRANS parser. Since most of the conceptual knowledge in the parser is contained in the conceptual knowledge base, which is separate from the parser's linguistic knowledge, this knowledge base could conceivably be used in other tasks, also. Thus, any new world knowledge learned by the parser would be available for other tasks using this knowledge base.

6. Using Generalized Syntactic Knowledge in an Integrated Parser

6.1 Introduction

The implementation of syntactic knowledge in terms of lexically-based requests was lacking in two respects: first, requests only performed "local" syntactic checks, and did not keep track of the parser's syntactic state. This lack of syntax-checking made it difficult to handle complex syntactic constructions without requiring a very large number of requests. Second, the integration of syntax and semantics in requests was so complete that general syntactic rules, such as a rule about the position and function of a verb's subject, were not expressible except by duplicating this information in the dictionary entries of every verb.

In this chapter, I will discuss a different approach to syntactic knowledge which does not use lexically-based rules, in contrast to many previous conceptual analyzers. This approach uses more autonomous syntactic knowledge, which is integrated dynamically with semantics during processing. Thus, the predictive advantages of integrated parsing are retained, while syntactic knowledge can be represented at the right level of generality. The approach is implemented in the MOPTRANS parser.

This approach also allows for more extensive building of syntactic representations during the parsing process, so that more global syntactic information can be used in order to help build a conceptual representation. Thus, the more extensive syntactic analyses required by complex syntactic constructions such as those I presented in chapter 4 can be accommodated in a more natural way than with lexically-based requests.

Finally, knowledge applicable to many languages need not be duplicated with this approach to syntax. Commonalities in syntactic construction among the languages that the MOPTRANS parser can parse, such as the fact that English and most romance languages are SVO languages, are reflected in the use of some of the same syntactic rules in these languages. Also, words which correspond to each other, such as "shoot" in English and "disparar" in Spanish, have identical lexical entries in MOPTRANS, thus reflecting their similarities to each other, and cutting down on the amount of duplication of knowledge in the system.

6.2 Generalizing Lexically-based Requests

Recall from chapter 4 the discussion about requests from the Conceptual Analyzer (Birnbaum and Selfridge, 1979) which looked for subjects of verbs. Almost all verbs in CA had some sort of request looking for a noun group to the left of the verb, which would fill some particular slot in the verb's conceptualization. These requests were all quite similar to each other, in that the same restrictions always applied to where the noun group could be, and what was done with the noun group was always the same. I demonstrated that the similarities among these requests could be abstracted out, to form

a single general request that could apply to all verbs:

> Subject request: Look back for a noun group which is not attached syntactically to anything before it. This noun group fills a particular slot (ACTOR, by default) in the conceptualization built by the word which activated this request. The word which activated this request will provide the name of the slot which should be filled, if it is not the ACTOR slot. The conceptualization built by the activating word will provide semantic restrictions on the noun group to be chosen by this request.

In this more general request, individual verbs, as well as the concepts built by these verbs, provide the information that is lost in the process of abstracting out common information from the original lexically-based requests. For example, here is the verb-specific information for a few verbs, along with the information provided by the concepts that they build:

> "Gave" information: "Gave" builds the conceptualization ATRANS.
>
> "Ate" information: "Ate" builds the conceptualization INGEST.
>
> "Received" information: "Received" builds the conceptualization ATRANS. The slot to be filled by the subject request is RECIPIENT.
>
> "Talked" information: "Talked" builds the conceptualization MTRANS.
>
> ATRANS information: The ACTOR and RECIPIENT of an ATRANS are ANIMATE.
>
> INGEST information: The ACTOR of an INGEST is ANIMATE.
>
> MTRANS information: The ACTOR of an MTRANS is a PERSON.

The general subject request can be rewritten in the following way:

> Subject rule: A noun group, which is not attached syntactically to anything before it, followed by an active verb, can be assigned as the subject of that verb. When this syntactic assignment is made, the representation of the noun group should be placed in a particular slot (ACTOR, by default) in the conceptualization built by the verb. The conceptualization built by the verb provides semantic restrictions on the noun group to be chosen by this rule.

Now the request has been turned into a declarative statement about one way in which a noun group and a verb can be combined. This rule provides information as to what this syntactic construction means semantically (namely, that the noun group will fill the ACTOR slot of the verb, or some other slot if the verb specifies). Thus, since there are pointers in the rule to semantic information that will be provided by a particular verb, all the semantic restrictions of the lexically-based requests above are still preserved.

This rule refers to purely syntactic concepts, such as "noun group" and "verb." Now that the particulars of each rule above have been abstracted out, such as the particular verb that activated the lexically-based requests, and the semantic restrictions on the conceptualization to the left of the verb, we are left with these purely syntactic concepts in the rule.

Much of the syntactic knowledge in the MOPTRANS parser is represented using

Generalized Syntactic Rules such as the one above. MOPTRANS still uses some lexically-based syntactic knowledge. For example, the fact that the word "for" follows the verb "to search" and indicates the OBJECT of the searching is encoded as a syntactic rule in the dictionary definition of "search." However, a large amount of syntactic knowledge is more appropriately represented on the level of syntactic categories, such as "verb," "subject," etc.; and is thus represented in MOPTRANS with Generalized Syntactic Rules. With syntactic knowledge expressed at the level of syntactic categories rather than at the level of individual words, the duplication of knowledge which was discussed in chapter 4 and the inability to share knowledge between languages is avoided.

Generalized Syntactic Rules in MOPTRANS have five parts to them. First, a rule contains a *syntactic pattern*, or a sequence of syntactic classes that must be found in active memory in order for the rule to apply. In the case of the Subject Rule above, the syntactic pattern is the appearance of a noun group followed by an active verb. Second, a rule can have a *syntactic assignment*, which indicates what syntactic role the elements in the rule play with respect to each other. In the Subject Rule, the noun group is assigned to be the subject of the verb, and a subject pointer is placed on the verb, pointing to the noun group. Third, a rule can have *additional restrictions*, which tells the parser other conditions under which the rule can or cannot apply. In this case, an additional restriction is that the noun group cannot be attached syntactically to anything before it. Fourth, a rule can have a *semantic action*, usually some slot-filling or concept-building action. In the Subject Rule, the semantic action is the filling of the ACTOR slot of the verb's representation with the noun group. Finally, a rule has a *result*, which specifies which elements in the rule remain in active memory, and what syntactic class these remaining elements now belong to. In the case of the Subject Rule, only the verb remains in active memory, because in general the subject will not be used in the course of building the representation of the remainder of the sentence[1]. The verb is also changed to the syntactic category S, indicating that it already has been assigned a subject.

In terms of these five features of Generalized Syntactic Rules, then, the Subject Rule consists of the following:

```
Subject Rule

Syntactic pattern:        NP, V (active)
Additional restrictions:  NP is not already attached syntactically
Syntactic assignment:     NP is SUBJECT of V, V is a MAIN CLAUSE
Semantic action:          NP is ACTOR of V (or another slot, if specified
                            by V)
Result:                   V (changed to S)
```

Let us examine some other requests in past conceptual analyzers, and their corresponding Generalized Syntactic Rules in the MOPTRANS parser. The word "gave," in past conceptually-based parsers such as CA, in addition to a request looking for a noun group to the verb's left to fill the ACTOR slot, also had lexically-based requests looking

[1]This is true with respect to the attachment of prepositional phrases, adjectives, etc., which occur later in the sentence. For example, in "The man asked the woman with glasses for a dime," it cannot be the case that it is the man who is wearing glasses, because "with glasses" appears after the verb. However, in the case of conjunctions, etc., the subject can be further used in the sentence. In cases like these, the pointer from the verb to its subject is used. This will be discussed later on.

for fillers of the OBJECT and RECIPIENT slots after the verb. These requests were the following:

> "Gave" OBJECT request: Look to the right of the verb for a noun group which has the property PHYSICAL-OBJECT, which is not attached syntactically to anything before it. Place the conceptualization in the OBJECT slot of the ATRANS.
>
> "Gave" RECIPIENT request: Look to the right of the verb for a noun group which has the property PERSON, which is either not attached syntactically to anything before it, or which is attached to the preposition "to." Place the conceptualization in the RECIPIENT slot of the ATRANS.

The "gave" OBJECT rule can be generalized with other similar requests for other transitive verbs, to form the following Generalized Syntactic Rule:

```
Object Rule

Syntactic pattern:        S, NP
Additional restrictions:  NP is not attached syntactically
Syntactic assignment:     NP is (syntactic) OBJECT of S
Semantic action:          NP is (semantic) OBJECT of S (or another
                            slot, if specified by S)
Result:                   S, NP
```

The syntactic pattern consists of an S followed by an NP because the Subject Rule changes the syntactic category of the verb (V) to an S. In the result, the NP is left in active memory, because prepositional phrases, etc., following the NP can modify either it or the S. (e.g., "The boy ate the cake with chocolate frosting," vs. "The boy ate the cake with a fork.")

The RECIPIENT request above for "gave" reflects the fact that "gave" is a verb which allows dative movement; that is, its indirect object can either appear after the preposition "to," or before the direct object. This request cannot be generalized for all English verbs, since only certain verbs allow dative movement. However, we can generalize the request, and others like it from other verbs, in terms of the following two rules:

```
Indirect Object Rule

Syntactic pattern:        S, PP
Additional restrictions:  PP begins with "to"
Syntactic assignment:     NP is (syntactic) INDIRECT OBJECT of S
Semantic action:          NP is (semantic) RECIPIENT of S (or another
                            slot, if specified by S)
Result:                   S, NP in PP
```

Dative Movement Rule

Syntactic pattern:	S, NP
Additional restrictions:	S has no syntactic OBJECT, NP is not attached syntactically, S allows dative movement
Syntactic assignment:	NP is (syntactic) INDIRECT OBJECT of S
Semantic action:	NP is (semantic) RECIPIENT of S (or another slot, if specified by S)
Result:	S, NP

These two rules express the fact that, for all verbs, the indirect object can be expressed after the preposition "to," and that for some verbs, which will be marked as allowing dative movement, the indirect object can be expressed as a noun group directly after the verb.

6.3 Integrated Parsing With Generalized Syntactic Rules

One of the goals in previous integrated parsers has been to combine syntactic and semantic processing, so that there are no separate stages of parsing. This facilitates the use of semantics to predict or disambiguate syntactic constructions, which is necessary in sentences such as the examples I presented in chapter 1, and desirable in general because it cuts down on the amount of incorrect syntactic decisions that are made. The use of lexically-based requests easily lent itself to this integration of process, due to the complete integration of syntactic and semantic knowledge. For example, consider the following sentence:

John gave Mary a book.

Because of the semantic information in the lexically-based requests which looked for the OBJECT and RECIPIENT of "gave," parsers like ELI and CA immediately assigned "Mary" to be the RECIPIENT of the ATRANS built by "gave," rather than the OBJECT, despite the syntactic ambiguity of the sentence at this point. This was because "Mary" fit the prototype of what should fill the RECIPIENT slot of an ATRANS better than the prototype of the OBJECT slot. This semantic information was reflected in the requests by the sorts of semantic objects which they looked for.

With Generalized Syntactic Rules, one must be more careful in the way in which the rules are applied and indexed in order to preserve the predictive and disambiguative power that integrated parsing provides. As we will see, the most straightforward way to apply these rules does not preserve this power. Thus, the MOPTRANS parser uses a more sophisticated indexing and application scheme for Generalized Syntactic Rules to achieve integrated application of syntactic and semantic knowledge.

A straightforward way to apply these rules would be to simply look for the appropriate syntactic patterns in active memory. In this approach, if a rule's pattern were matched, then the rule would be executed, provided the elements matching the pattern were semantically appropriate for performing the semantic action of the rule. To explain what I mean by "semantically appropriate," consider the following two sentences:

The man wrapped the present.
The present wrapped by the man was expensive.

The syntactic patter of the Subject Rule would match in both of these senter However, the Subject Rule should not be executed in the second sentence, bec: "present" does not meet the prototype of the ACTOR of "wrapped," because it does refer to a PERSON. Thus, in the second sentence, "the present" would not semantically appropriate for the Subject Rule, and the rule would not be executed.

The simple rule application scheme, then, would be for the parser to look syntactic patterns in active memory corresponding to patterns in its Generalized Synt: Rules. If a rule matched, and if the elements matching the rule were semanti appropriate, the semantic action of the rule would be executed, and active memory w be modified according to the RESULT property of the rule.

In cases where more than one rule could apply at once, rules would have t prioritized. For example, a noun group followed by a verb which could either be active or past participle would match the syntactic patterns of both the Subject Rule an Unmarked Passive Rule, which would look for an NP followed by a past participl the NP and the verb were semantically appropriate for both rules, then the parser w not know which rule to apply unless one rule had priority over the other. Thus, in straightforward scheme we would want to give priority to the "more basic" rules, so the parser would favor actives over passives, etc., in cases where both were pos semantically.

This sort of simple application of Generalized Syntactic Rules would preserve son the advantages of integrated parsing. For example, syntactic constructions that did make sense semantically would not be pursued. Thus, an *irreversible passive*, suc "The present wrapped ..." in the example above, would be parsed immediately passive, rather than considering the active construction only to find syntactic cues lat the sentence indicating unmarked passive.

However, some of the power of integration would be lost in this scheme. This w be true whenever two possible interpretations of a sentence existed, but semantics str preferred one interpretation over the other, even though both were semantically plau: For instance, in "John gave Mary a book," an ambiguity exists after reading "M: since "Mary" could be the OBJECT of the ATRANS, rather than the RECIPIE However, since "Mary" fits the prototype of the RECIPIENT slot much better tha OBJECT slot, it makes sense to choose the RECIPIENT interpretation over the OBJ interpretation. This would not occur using Generalized Syntactic Rules in the simple I have outlined. Presumably, the Object Rule would be given preference over the D Movement Rule. Thus, the parser would first choose the interpretation that "Mary" the OBJECT of the ATRANS, since this is semantically acceptable, even thoug other interpretation is certainly preferable, and in this case turns out to be right. applying Generalized Syntactic Rules in this way would result in the parser havi back up in cases where it does not seem that it should have to.

In order to preserve the ability to choose semantically preferable interpretatio syntactically ambiguous constructions, which is one of the main advantages of integ parsing, the MOPTRANS parser indexes Generalized Syntactic Rules according to semantic actions, in addition to their syntactic patterns. To choose a rule to be exe the MOPTRANS parser examines all the conceptualizations in active memory. It tr find connections between these conceptualizations; that is, it tries to find a slot i conceptualization into which another conceptualization will fit. Once it has foun

possible connections between the elements in active memory, it selects the connection which is "best"; i.e., the one in which the potential slot-filler meets most closely the prototype for what should fill that slot. After it has selected the best connection, it looks for a Generalized Syntactic Rule whose semantic action will perform that connection. If it finds such a rule, and the elements which it wants to connect also meet the syntactic pattern of the rule, then the rule is performed. Otherwise, the parser chooses the next-best connection, and looks for a rule to perform this slot-filling. This continues until either a rule is executed, or no more connections are left. If this process fails to find a rule to be executed, then the parser finds a rule according to the syntactic indexing method discussed above. This rule selection process is displayed graphically in Figure 6-1.

To make this more clear, consider how the MOPTRANS parser selects Generalized Syntactic Rules for the examples which I discussed above. In "John gave Mary a book," after reading the word "Mary," the parser's active memory contains the representation of "gave," (ATRANS ACTOR (HUMAN GENDER MALE NAME JOHN)), along with the information that this representation is currently classified syntactically as an S; and the representation of "Mary," (HUMAN GENDER FEMALE NAME MARY), along with the information that this is an NP (how the parser labels this as an NP will be discussed in detail in chapter 7). "John" is no longer in active memory, because the Subject Rule has removed it. In beginning to select a rule at this point, MOPTRANS considers what connections could be made between the ATRANS and the HUMAN. It finds two possible connections: that the HUMAN is either the RECIPIENT, or the OBJECT of the ATRANS (A HUMAN could also be the ACTOR or the SOURCE (FROM) of an ATRANS, but these slots are already filled). Because a HUMAN meets the prototypes of the RECIPIENT and SOURCE slots of the ATRANS better than the OBJECT slot, these are the two connections which the parser would prefer. Since there is no preference between these two, it arbitrarily picks one for which to find a Generalized Syntactic Rule. Among the rules which would perform these slot-fillings is the Dative Movement Rule, which fills the RECIPIENT slot of the ATRANS (the Object Rule is NOT one of the rules which is found). This rule is the only rule which the parser finds whose syntactic pattern is matched by active memory. Thus, the Dative Movement Rule is chosen, and "Mary" is assigned to be the RECIPIENT of the ATRANS.

Now, consider the active vs. passive examples from before:

The man wrapped the present.
The present wrapped by the man was expensive.

The semantic indexing scheme selects the appropriate rules to be executed in both of these examples, also. When the MOPTRANS parser encounters "wrapped" in the first example, active memory contains (HUMAN GENDER MALE), categorized as an NP; and a representation for "wrapped," say COVER, categorized as either a verb (V) or a past participle verb (VPP). Only one possible connection is found at this point, that the HUMAN is the ACTOR of the COVER (conceivably, a HUMAN could be the OBJECT of a COVER, also, but HUMAN better matches the prototype for the ACTOR slot). This slot-filling can be accomplished by only one rule which matches current syntactic conditions, the Subject Rule, which is chosen to be executed.

In the second example, "present" is in active memory instead of "man." This time, the only possible connection found is that "present" can be the OBJECT of COVER. Among the rules which can perform this slot-filling is the Unmarked Passive Rule, which is the following:

```
                        ----------------------------
                       | Find possible connections  |
                       | between concepts in active |
                       | memory                     |
                        ----------------------------
       -------------------->|
      |                     |
      |                    / \
      |                   / Are \
      |                  /  there  \
     /|\                /  possible  \ ---no-------------
      |                 \   connec-   /                  |
      |                  \  tions    /                   |
      |                   \   ?     /                    |
      |                     \ /                          |
     /|\                     |                           |
      |                     yes                          |
      |                      |                           |
      |              -------------------------           |
      |             | Choose "best" connection |         |
     /|\             -------------------------           |
      |                      |                           |
      |           ------------------------------         |
      |          | Find Generalized Syntactic   |        |
      |          | Rules whose semantic actions |        |
     /|\         | will make this connection    |        |
      |           ------------------------------    ---------------
      |                      |                     | Select a rule |
      |           ------------------------------   | by indexing   |
      |          | Choose a rule whose syntactic | | via syntactic |
     /|\         | pattern is also satisfied     | | patterns      |
      |           ------------------------------    ---------------
      |                      |                            |
      |                     / \                           |
      |                    /   \                          |
 -----------------        /Is   \                         |
| Remove the      |      /       \                 ------------------
| connection from |<--no-- / there \ --yes------> | Execute the rule |
| the list of     |       \ such a/                ------------------
| possible        |        \rule?/
| connections     |         \   /
 -----------------           \ /
```

Figure 6-9: Generalized Syntactic Rule Selection Process in the MOPTRANS Parser

Unmarked Passive Rule

```
Syntactic pattern:         NP, VPP
Additional restrictions:   none
Syntactic assignment:      NP is (syntactic) SUBJECT of VPP, VPP is PASSIVE,
                              VPP is a RELATIVE CLAUSE of NP
Semantic action:           NP is (semantic) OBJECT of S (or another
                              slot, if specified by VPP)
Result:                    NP, VPP (changed to S)
```

This rule has some differences from the Subject Rule. First, there is no additional restriction that the NP not be syntactically attached to anything before it, since unmarked relative clauses can be attached to any NP. Second, the NP is left in active memory, since the rest of the sentence after the clause can refer back to the NP, unlike NP's used as subjects.

Since the Unmarked Passive Rule is the only rule which both makes the desired connection and matches syntactic conditions of active memory, it is selected to be executed. Thus, MOPTRANS assigns "present" to be the OBJECT of "wrapped" in this example.

The simpler, syntactic indexing scheme I discussed earlier was also able to choose the correct rule to be run in these two examples. However, in this scheme, it was necessary for the parser to consider both the Subject Rule and the Unmarked Passive Rule for both sentences. In the first example, the Unmarked Passive Rule was considered, since it matched the syntactic pattern in active memory, but the Subject Rule was chosen because it had priority over this rule. In the second example, the Subject Rule was again chosen. Only after consideration of semantics was this rule then discarded and the Unmarked Passive Rule chosen.

With the semantic indexing scheme used in MOPTRANS, the correct rule is chosen in these two examples without the parser even considering the incorrect rule. This is because semantics guides the rule selection process instead of syntax. Because of this, Generalized Syntactic Rules which would produce semantically incoherent representations are not even considered by the parser when selecting the next rule to be executed.

6.4 Correcting Errors

Although the semantic rule-indexing system used in the MOPTRANS parser usually chooses the correct interpretation of an ambiguous sentence, there are times when the wrong rule is chosen, because the sentence is semantically misleading. For instance, consider the following example, presented in (Wilks, 1975b):

John gave Mary to the Sheik of Abracadabra.

In this sentence, the MOPTRANS parser, and presumably most human readers, must correct an assumption that it has made, that "Mary" is the RECIPIENT of "gave."

To handle situations in which it has made the wrong inference, the MOPTRANS parser uses two types of rules: error correction rules and backup rules. Both are indexed strictly by syntactic patterns, as opposed to most of the other parsing rules, which are indexed by semantic actions and syntactic patterns. These rules identify situations,

according to syntactic patterns in active memory, in which the wrong Generalized Syntactic Rule has been applied sometime during the parse of a sentence.

An example of an error correction rule is the rule which would correct the error made in the Sheik of Abracadabra example above:

```
Error Correction Rule 1:

Syntactic pattern:         S, PP
Additional restrictions:   S has had Dative Movement Rule run on it,
                           PP starts with "to"
Syntactic assignment:      NP inside the PP is the INDIRECT OBJECT of S,
                              old NP formerly assigned as the INDIRECT
                              OBJECT of S is the (syntactic) OBJECT of S.
Semantic action:           NP inside of PP is the RECIPIENT of S, old NP
                              formerly assigned as the RECIPIENT of S
                              is the (semantic) OBJECT of S.
Result:                    S, NP inside PP
```

This rule is similar in spirit to the rule proposed in (Birnbaum and Selfridge, 1979) for this example, although their request-based version of this rule was much more limited, in that it only applied to active sentences involving the word "gave."

Error correction rules are indexed only according to their syntactic patterns because of the restriction in the semantic indexing procedure that only unfilled slots are considered as possible connections that should be made between elements in active memory. Thus, in this case, although "Sheik of Abracadabra" would fit well into the RECIPIENT of the ATRANS, semantics would not index to Error Correction Rule 1 because of the fact that this slot is already filled.

In other cases, when it can be identified that a mistake has been made, it is not easy for the parser to know what sort of correction to make. In cases like these, the MOPTRANS parser uses backup rules. A backup rule is used during the parse of the following example, discussed in (Marcus, 1978):

The horse raced past the barn fell.

The verb "race" has a slightly different meaning when used transitively and intransitively. In both cases, the word refers to the primitive action PTRANS, with the stipulation that the SPEED of the PTRANS is RAPID. However, in the transitive case, the ACTOR and OBJECT of the PTRANS are different, whereas in the intransitive case, these slots are filled by the same conceptualization (i.e., "The horse raced past the barn" is equivalent to "The horse raced *himself* past the barn").

Because this is the case for some past active / past participle verbs, it is difficult to use error correction rules when the parser has inferred the wrong syntactic role for these verbs, since a different semantic representation must be built, in addition to the corrections that must be made. Thus, a backup rule is used instead. Instead of correcting the mistake at the time of its identification, as with error correction rules, the parser undoes the parsing which has taken place since the error was made. The backup rule for this situation is the following:

Backup Rule 1:

```
Syntactic pattern:        S, V
Additional restrictions:  S is a MAIN CLAUSE
Action:                   Back up to the execution of the Subject Rule
                            on the S.
```

In the parsing of "The horse raced past the barn fell," the MOPTRANS parser first assumes that "raced" is an active verb. This is because "horse" fits better into the ACTOR slot of PTRANS, since PTRANS requires an ANIMATE ACTOR but any PHYSICAL OBJECT as its OBJECT. This assignment seems fine until the parser encounters "fell." At this point, no Generalized Syntactic Rules can attach "fell" to anything before it in the sentence. Semantically, "the barn" could be the OBJECT of "fell" (which is a PTRANS). However, no Generalized Syntactic Rules have syntactic patterns which allow this slot-filling to take place. Because no rules can be executed, Backup Rule 1 applies, and the parser backs up to the state it was in when it executed the Subject Rule. This rule is prohibited from being executed, and thus the Unmarked Passive Rule is chosen instead. Then, parsing of the remainder of the sentence proceeds smoothly, since "the horse" is still on the active list to combine with "fell," using the Subject Rule.

Since the MOPTRANS parser uses backup rules such as the one above, it is sometimes necessary for the parser to remember parsing states as it proceeds through a sentence. This may seem like a large burden to place on the parser. However, the number of situations in which the parser is required to remember its state has turned out to be fairly small. Certain rules are marked as to whether or not the parser should remember its state before executing the rule, and when this is necessary. The Subject Rule is one such rule. It is marked so that the state of the parser is saved whenever the verb is intransitive and could also be a past participle verb.

In the sentences that the MOPTRANS parser has encountered, the number of rules which require that the state of the parser be saved is relatively small. Even then, these rules do not always require that the parser's state be saved. For instance, the Subject Rule does not require a state save when transitive active verbs or active verbs that could not be past participles are involved. Thus, the amount of extra work required to save parsing states has proven to be minimal. The number of situations in which backup is necessary has been minimal, also. Moreover, these situations seem to correspond to garden path sentences, in which people would presumably be misled and forced to reparse the sentence. "The horse raced past the barn fell," is an example of a garden path sentence.

6.5 Generalized Syntactic Rules, Complex Syntactic Constructions, and Syntactic Ambiguities

The use of Generalized Syntactic Rules to parse complex syntactic constructions or sentences containing syntactic ambiguities results in the need for far fewer parsing rules than were needed with lexically-based requests. Consider the examples from chapter 4 illustrating the use of verbs which could function as either past active or past participle:

Example 1: The soldier called to the sergeant shot in the arm.

Example 2: The soldier called to the sergeant shot three enemy troops.

Two sets of requests were needed to disambiguate verbs such as "called." One set looked for cues such as the appearance of a form of "to be" to the left of the verb in question, or the presence or absence of another active verb in the sentence. The other set of requests was for the special case in which another verb which could either be past active or past participle was found in the sentence. In the above examples, this second set of requests was needed to determine first if "shot" was active or passive, which would then in turn determine if "called" was active or passive. A total of eight requests were needed for the verb "called," and it was evident that similar numbers of requests would be needed for all verbs which could either be past active or past participle.

Except for the request which looked for a form of "to be," all of the other requests needed were in essence looking for another verb in the sentence which could function as the main verb. If another such verb was found, then the verb in question was a past participle. If no main verb was found elsewhere in the sentence, then the verb in question was past active.

The reason that so many requests were required was that the parser was not normally keeping track of whether or not the main verb of the sentence had been encountered. Unfortunately, it is not easy to tell if a verb is the main verb of the sentence unless syntactic processing has been going on throughout the parsing of the sentence. Unless verbs are marked as main verbs or dependent clause verbs during the course of normal processing, it is hard to look at a particular verb in a sentence and determine on the fly whether or not that verb is the main verb. This was the reason that so many requests were needed: the task of determining syntactic functions of words, such as whether or not a given verb in a sentence is the main verb, requires examining a great deal of the surrounding syntactic context.

In the MOPTRANS parser, since verbs are marked during the normal course of parsing as to what syntactic function they are serving, the rules needed by the parser to disambiguate examples such as the ones above are much simpler than the requests which were needed. Example 1 is parsed correctly using the Subject Rule and Unmarked Passive Rule. When the parser reads "called," it finds two possible connections between "soldier" and the MTRANS representing "called": the soldier could either be the ACTOR or the RECIPIENT of the MTRANS. The parser has no preference between these two possible slot-fillings, since they both fit the prototypes of the slots equally well. Thus, the Subject Rule is selected, since it has preference over the Unmarked Passive Rule in cases where there is no semantic preference. When the parser reads "shot," again it finds two possible connections: "sergeant" could either be the ACTOR or the OBJECT of the concept SHOOT. This time, only the OBJECT slot-filling can be performed by the Generalized Syntactic Rules, since the Unmarked Passive Rule is the only one that applies. (The Subject Rule does not apply, because "sergeant" is already attached syntactically, since it is the syntactic INDIRECT OBJECT of "called.")

Example 2 requires an additional backup rule. The parsing of this sentence proceeds in exactly the same manner as for example 1, until the parser reads the NP "three enemy troops." This NP cannot be attached to anything, since "shot" does not expect a direct object, because it is marked as passive. Thus, the following backup rule is executed:

Backup Rule 2

```
Syntactic pattern:          S, NP
Additional restrictions:    S is a RELATIVE CLAUSE, S is UNMARKED PASSIVE
Action:                     Back up to the execution of the Unmarked Passive
                               Rule on the S.
```

When this backup rule is executed, the parser returns to the state it was in before "shot" was assigned to be an unmarked passive verb. The Unmarked Passive Rule is prohibited from executing again at this point. But since the Unmarked Passive Rule was the only rule that could be executed, the parser now executes Backup Rule 1. This causes the parser to back up further, undoing the assignment of "soldier" as the ACTOR of the MTRANS. The Subject Rule is prohibited from executing, and so the parser selects the Unmarked Passive Rule, assigning "soldier" to be the RECIPIENT of the MTRANS. The remainder of the sentence is then reparsed. When the parser reads "shot" the second time around, it finds four possible connections: the soldier could be the ACTOR or the OBJECT of the SHOOT, or the sergeant could fill either of these slots. (The first time around, the first two connections were not possible, because "soldier" had been removed from active memory by the Subject Rule. However, this time, the Unmarked Passive Rule has left "soldier" in active memory.) Two Generalized Syntactic Rules could perform two of these slot-fillings: the Subject Rule could assign the soldier to be the ACTOR of SHOOT, or the Unmarked Passive Rule could assign the sergeant to be the OBJECT of SHOOT. Since there is no semantic preference between these two connections, the parser selects the Subject Rule, assigning "soldier" as the subject of "shot" and "shot" as the main verb of the sentence. The parse of the remainder of the sentence proceeds smoothly.

These rules also handle even more complex sentences, such as the following:

Example 3: The soldier called to the sergeant shot in the arm was reprimanded.

This example, which caused problems for the lexically-based requests, can also be handled by the rules which I have presented so far. Although this is a difficult sentence for people to understand, and is not a typical sentence, it demonstrates the robustness of MOPTRANS' syntactic rules. MOPTRANS successfully parses this sentence as follows: first, "soldier" is assigned to be the ACTOR of "called." Then, "shot" is assigned as an unmarked passive, with "sergeant" as the OBJECT of SHOOT. Parsing continues, until "was reprimanded" is read. The Passive Rule assigns "reprimanded" to be a V. At this point, no rules can attach "reprimanded." Thus, Backup Rule 1 applies, since a MAIN CLAUSE verb is followed by another V. This causes the parser to back up to the initial assignment of "soldier" as the subject of "called." The second time through, "soldier" is assigned as the OBJECT of the MTRANS. But "shot" is chosen as the main verb, since "soldier" fits as the ACTOR of SHOOT. Again, Backup Rule 1 applies when the parser reads "was reprimanded." This time, the assignment of "soldier" as the subject of "shot" is undone, "shot" is assigned as an unmarked passive, with "sergeant" as the OBJECT of SHOOT, and "was reprimanded" is finally assigned as the MAIN VERB of the sentence.

Thus, we see that due to the explicit assignment of verbs' functions with this set of rules, the rules required to disambiguate this class of verbs, even in very syntactically complex sentences, are simple and straightforward. The number of rules required is much smaller than was the case with lexically-based requests which did not compute the syntactic functions of verbs, and more complex examples, such as "The soldier called to

the sergeant shot in the arm was reprimanded," can be handled.

Although the number of rules needed by the parser is small, this does not mean that the parsing process during the parse of sentences such as example 3 above is simple. Two backups must be performed by the MOPTRANS parser to finally understand this example correctly. This seems to parallel problems encountered by human readers in examples such as this. Although people eventually understand syntactically complex sentences such as example 3, it is not without difficulty and one or more re-readings. Thus, the MOPTRANS parser seems to parallel the same process that human readers must go through in order to parse such sentences.

6.6 Comparison to Syntactic Parsers

Given that syntactic knowledge in the MOPTRANS parser is represented more autonomously than in previous conceptual analyzers, it is interesting to compare the Generalized Syntactic Rules in MOPTRANS to the parsing rules used in syntactic parsers. These rules are similar in some ways to the parsing rules used in some syntactic parsers, most notably PARSIFAL (Marcus, 1978). In PARSIFAL, syntactic rules were applied as the text was parsed in a left-to-right manner, just as in MOPTRANS. The syntactic rules in PARSIFAL took on the same basic form, looking for patterns of syntactic constituents in the input text. For example, the equivalent rule in PARSIFAL to the Subject Rule was the following:

RULE UNMARKED-ORDER IN PARSE-SUBJ:
[=np] [=verb] --> Attach 1st to S-node as NP, Deactivate PARSE-SUBJ,
    Activate PARSE-AUX

Parsing rules in PARSIFAL were members of "packets," which were activated and deactivated during the parsing process. Thus, this rule was in the packet PARSE-SUBJ, responsible for finding the subject of a sentence. This rule looked for an NP followed by a V in PARSIFAL's input buffers, and assigned the NP as the subject of the sentence if this pattern was found. In addition, a new packet of rules, PARSE-AUX, was activated to parse any auxiliary verbs occurring before the main verb.

One of the motivations behind PARSIFAL was to avoid the processing inefficiencies which were apparent in ATN parsers, such as LUNAR (Woods, Kaplan and Nash-Webber, 1972), whose top-down nature caused them to back up excessively. Although an ATN's parsing rules are based on similar syntactic patterns, they are applied in a more top-down way. Thus, many hypotheses created by this top-down application are immediately rejected by the input. For example, consider the following syntax rules:

S -> NP VP
S -> PP S

An ATN with these grammar rules would immediately push for an NP, assuming that the first rule had priority over the second rule. If the initial word in the sentence turned out to be a preposition, a backup would be required, even though no processing of the input had even taken place so far.

As Marcus also pointed out, excessive backup was required in ATN's in sentences like the following:

Is the block sitting in the box red?
Is the block sitting in the box?

In case like these, when the discovery that backup is required is made long after the application of the wrong rule, an ATN parser must do a lot of backtracking, since it does not know what rule was misapplied. It will take awhile before the misapplied rule will be found, since it was executed so long ago in the sentence.

The limited backup mechanisms in the MOPTRANS parser are motivated for similar reasons. Although the Generalized Syntactic Rules could be applied in the same top-down manner as in ATN's, with automatic backtracking whenever a new lexical item entered that did not match the pattern of any rule, the MOPTRANS parser's rules are applied in a more bottom-up way. Thus, the parser does not need to immediately account for the appearance of every new lexical item, but rather can wait for the building of a larger syntactic constituent before a rule matches. In this way, the control structure of the MOPTRANS parser is similar to PARSIFAL's.

Aside from these similarities, the differences in goals between a syntactic parser like PARSIFAL and the MOPTRANS parser are substantial. First, since a conceptual representation is built directly in the MOPTRANS parser, parsing rules do not correspond to transformational rules, as they do in PARSIFAL. With passive sentences, for instance, PARSIFAL builds a trace NP, which it places in the active memory buffer. The direct object rule then matches on this trace, assigning this trace as the direct object of the verb. The argument for doing this is to capture the linguistic similarities between active and passive sentences in the representation. In a conceptual analyzer such as MOPTRANS, however, there is no need to do this, since the similarity is already captured by virtue of the same conceptual representation which is built for actives and passives.

Another difference between MOPTRANS and PARSIFAL is the way in which rules are indexed. PARSIFAL uses the approach of matching syntactic patterns in active memory. This approach is not used in MOPTRANS, due to the objections which I presented earlier in this chapter, that this indexing system would not take advantage of the predictive power of integrated syntax and semantics.

Given MOPTRANS' indexing approach, it is possible to explain some garden path phenomena that cannot be explained by Marcus' theory of parsing. Marcus claimed that his parser was capable of parsing all sentences deterministically except for garden path sentences, such as "The horse raced past the barn fell." However, his syntax-first approach to parsing is not able to explain why some sentences are garden path, while other sentences with exactly the same syntactic construction are not garden paths. For example, here are two such sentences, which were presented in (Crain and Steedman, 1982):

The teachers taught by the Berlitz method passed the course.
The children taught by the Berlitz method passed the course.

Crain and Steedman reported that subjects experienced more difficulty interpreting the first example above than the second example, indicating that the first example was indeed a garden path, but the second example was not. Presumably this was due to the semantic preference of "teachers" as the AGENT of "taught," but "children" as the PATIENT of "taught."

Since only semantic/pragmatic considerations can explain why the first of these sentences is a garden path, but the second sentence is not, Marcus' parsing algorithm cannot account for this difference. However, because the MOPTRANS parser indexes

syntactic rules according to their semantic actions, the second sentence would not be a garden path for MOPTRANS. This is because the semantic indexing scheme would select the Unmarked Passive rule to be executed on "students" and "taught," since "students" fits better and the OBJECT of "taught" than as its ACTOR.

6.7 Processing Ungrammatical Sentences

One possible criticism of the use of Generalized Syntactic Rules is that it is not clear how these rules could process ungrammatical sentences. In the algorithm presented in Figure 6-9, any slot-filling action performed had to be performed by a Generalized Syntactic Rule. Since the execution of a rule required that the rule's syntactic pattern be matched, only grammatical patterns would result in a rule being executed.

One possible solution to this problem is the solution proposed by Charniak in the PARAGRAM parser (Charniak, 1981). PARAGRAM used PARSIFAL-like situation/action rules, which looked for syntactic patterns in the parser's input buffer. However, instead of using simple yes/no tests, the result of a test in PARAGRAM was a numerical "goodness rating." A better fit between the contents of the buffer and the rule's test resulted in a higher goodness rating.

Instead of testing parsing rules until one test was satisfied, as was the case in PARSIFAL, PARAGRAM evaluated the tests of many parsing rules, and chose the one with the highest goodness rating. Thus, even if the contents of the buffer did not match exactly with PARAGRAM's parsing rules, as would be the case with ungrammatical input, some rule was always chosen. As a result, the parser was able to parse examples of ungrammatical sentences.

A similar approach to parsing ungrammatical sentences could be employed in the MOPTRANS parser. Generalized Syntactic Rules could also be given "goodness ratings," so that the result of testing to see if a Generalized Syntactic Rule applied to a given situation would not be a simple yes/no. Then, if semantic indexing found no Generalized Syntactic Rules which exactly matched the syntactic pattern in active memory, the goodness ratings could be used to select a rule anyway. This would be done just as it was in PARAGRAM: the goodness ratings of all the Generalized Syntactic Rules which could possibly apply in the given situation could be computed, and then the rule with the highest rating would be chosen.

6.8 Generalized Syntactic Rules in a Multi-lingual Parser

Recall that with lexically-based requests, sharing knowledge across languages was virtually impossible. This was because of the high level of integration of knowledge in requests. Syntactic knowledge was mixed in with semantic knowledge, disambiguation knowledge, etc. Thus, for instance, the dictionary definition for a word like "shot" would be almost completely useless in writing the word definition for the equivalent verb in Spanish, "disparar," even though the use of this word in Spanish parallels for the most part its use in English. This is because the requests in the dictionary definition of "shot" are used to disambiguate between the past participle and past active uses of the verb, an ambiguity that does not occur in the Spanish. Thus, few of the requests in the dictionary

definition of "shot" would have any use for the word "disparar."

This is not the case, however, with Generalized Syntactic Rules. In the MOPTRANS parser, the dictionary definition of "shot" is very simple. It simply states that "shot" is either an active verb or a past participle, and that it builds a conceptual representation called SHOOT. All the other knowledge needed to parse this word is contained in the semantic knowledge that the parser has about the concept SHOOT -- that SHOOT takes an ACTOR who is HUMAN, an OBJECT which is a PHYSICAL-OBJECT, an INSTRUMENT which is a GUN, and the RESULT of SHOOT is often that the OBJECT is DAMAGED in some way -- and the syntactic knowledge that the parser has about past active and past participle verbs. Thus, the Spanish verb would have a nearly identical definition.

Individual syntactic rules can also be shared across languages. For example, Spanish, English, and French are all SVO languages. A noun group appearing before a verb which is not be attached syntactically to anything before it can function as the verb's subject, and fills a certain slot, ACTOR by default, of the conceptualization built by the verb. Thus, in the MOPTRANS parser, exactly the same subject rule is used for English, Spanish, and French. This was not possible with request-based syntax, since the individual rules in the dictionary entries of verbs often had other functions, such as the function of disambiguating the verb, in addition to the function of finding the subject of the verb.

6.9 Generalized Syntactic Rules and Learning

Although the task of learning to parse is beyond the scope of this thesis, it is important to examine the learnability of the syntactic rules which are used in the MOPTRANS parser. In chapter 4, I contended that it would be very difficult, if not impossible, to learn syntactic knowledge in the form of lexically-based requests, because of the lack of generality of these rules. For instance, consider requests which would be found in the dictionary entry for the word "shot." Some of these requests, such as the request looking for the preposition "in" after the verb, followed by a BODY-PART which would be the part of the victim's body that was wounded, are rather specific to the verb "shot," and do not apply to other verbs. However, other requests, such as those which determine whether "shot" is past active or past participle, could apply, with a small amount of modification, to a larger class of verbs, namely those verbs which can be either past actives or past participles. Finally, other information in the requests, such as the fact that when "shot" is active, the ACTOR of "shot" often appears to the left of the verb, and the OBJECT to the right, applies to verbs in general. However, nowhere in these requests is this stated. All of the requests are written specifically for the verb "shot."

The learning problem, then, is that when a new verb is learned, the learner cannot determine which requests that he knows from other verbs can apply to the new verb. Is it the case for the new verb that the preposition "in" will be followed by the HURT-PART slot of its conceptualization? Or should the learner even infer that the new verb builds the same conceptualization as "shot"? How about the rules which determine whether "shot" is past active or unmarked passive? Do these rules apply to the new verb? In short, since none of this knowledge is marked as to how general it is, a learner cannot infer whether or not any of it applies to a new verb just being learned. Since this is the case, this implies that the learner would have to learn *everything* about how this new verb

functions, including where to look in the sentence for the slot-fillers of its conceptualization, how to disambiguate it if it is ambiguous, what particular prepositions indicate particular slots, etc.

Clearly if nothing can be inferred about a new word from words that are already known, the task of learning an entire language would be hopelessly complex. A learner would be forced to learn an entirely new and intricate set of rules for every single word in the language. This is a ridiculously hopeless task, given the number of words in natural languages. So the lexically-based approach to syntactic knowledge appears to be incompatible with the task of learning a natural language.

This problem would not occur in a learner in which syntactic knowledge was represented as it is in the MOPTRANS parser. When learning a new word, the category to which the word belongs would provide a large amount of knowledge as to how to parse the new word. This is because syntactic knowledge is stored at the appropriate level of generality with this approach to syntax. A rule saying that "in" following the word "shot" could indicate that the hurt portion of a victim's body will follow the preposition is stored in the dictionary definition of "shot," whereas subject or unmarked passive rules are stored under the appropriate categories, V and VPP (past participle verb). Thus, when a new verb is learned, the appropriate rules would apply to the new word depending on the new word's syntactic category.

6.10 Implementation of Generalized Syntactic Rules

Generalized Syntactic Rules are implemented in the MOPTRANS parser in the form of production rules, which consist of two possible types of tests and an action. The action consists of some combination of semantic actions, such as a slot-filling or the merging of two conceptualizations; and syntactic actions, such as the assigning of a new syntactic category to one of the elements in active memory. These actions can also add or remove new elements to active memory.

Since I have argued that Generalized Syntactic Rules should be indexed in terms of their semantic actions, one type of test in these production rules consists of the semantic action that takes place during the execution of that rule, along with the syntactic types of the elements that the action should be performed on. For example, the Subject Rule for English is indexed according to the fact that it fills a particular slot of a verb with a noun phrase. The particular slot is normally the ACTOR slot, since by default this is the slot which the subject rule fills in, but the particular verb provides the indexing scheme with the slot that the Subject Rule should fill.

Generalized Syntactic Rules are also indexed according to the order of syntactic elements that should appear in active memory in order for a rule to be executed. Thus, the Subject Rule is also indexed by the appearance of a noun phrase followed by a verb in active memory. This double indexing is necessary because sometimes semantics does not have enough information to index to the right rule. This occurs generally for two classes of rules (although we will see shortly that syntactic indexing is necessary in other cases): rules which do not perform semantic actions, and rules which operate on conceptualizations which are not longer in active memory.

An example of a rule with no semantic action is the rule which finds the head noun of a noun group. Such a rule is needed in English to distinguish the use of nouns as adjectives from their use as nouns (e.g., "the Mexican restaurant" as opposed to "the

Mexican"). The need for such a rule in English is discussed in greater detail in the next chapter. The rule is as follows:

Head Noun Rule

Syntactic pattern:	N, <any word>
Additional restrictions:	The word following the N is not a N.
Syntactic assignment:	none
Semantic action:	none
Result:	N (changed to HN), <any word>

The only action performed by this rule is to change the syntactic category of the noun to a Head Noun (HN). Thus, it cannot be indexed according to its semantic action, since it has none. Because of this, the MOPTRANS parser uses the syntactic indexing methods discussed earlier in this chapter to find this rule.

Other rules, which do perform semantic actions, must also be indexed syntactically, because the conceptualizations on which they operate are not all in active memory. For example, here is one of the rules for attaching participial phrases following prepositions:

Participial Phrase Rule 1

Syntactic pattern:	S, PREP, V
Additional restrictions:	V is a present participle
Syntactic assignment:	V is a CLAUSE of S, SUBJECT of S is SUBJECT of V.
Semantic action:	SUBJECT of S is the ACTOR of V (or another slot, if specified by V), V fills semantic slot of S as specified by PREP.
Result:	S, V (changed to S)

This rule is used for sentences such as the following:

John read the book after borrowing it from Mary.

This rule assigns "John" as the ACTOR and RECIPIENT of the ATRANS representing "borrowing" (this verb specifies that its subject should fill both semantic slots), and also assigns the temporal relation AFTER between the actions READ and ATRANS.

Since this rule performs two semantic actions, it could conceivably be found by semantic indexing in two different ways: the parser could notice that "read" and "borrow" could have the semantic relation AFTER between them, or it could notice that "John" could be the ACTOR or RECIPIENT of the ATRANS representing "borrow." However, neither of these semantic connections is considered by the parser. This is because the relation AFTER could occur between any two actions, and thus there is no AFTER slot in the case frame of either READ or ATRANS. Thus, this possible connection is never found. The connection between "John" and the ATRANS is also never found, because the execution of the Subject Rule removes "John" from active memory. Thus, this rule is found by syntactic indexing.

To some extent, the fact that rules such as the Participial Phrase Rule above cannot be indexed semantically violates the desire to keep processing as integrated as possible. Since semantics cannot predict this syntactic structure, the parser must occasionally attempt to execute this syntactic rule, even though semantically it does not make sense.

However, it is important to constrain the amount of search that the parser must p in order to find potential semantic connections. Otherwise, the benefits of inte parsing, brought on by the predictive power of semantics, would be lost, because t of this predictive power would outweigh the gains brought on by not pu conceptually senseless syntactic constructions. Thus, the indexing technique used MOPTRANS parser is an attempt to find a good compromise between using as semantics as early as possible in the parsing process, without too great a search cost

6.11 Rule Application and Semantic Failures

The rule selection process in the MOPTRANS parser was illustrated in Figu Once a rule is selected by this algorithm, the rule is executed. During the co executing a rule, it is possible that a "semantic failure" may occur. A failure occur a semantic prototype is violated during a slot-filling, or some other semantically action is performed. If this happens, the state of active memory is returned to w was before the execution of the rule, and the rule selection process is repeated t another rule to be executed. This continues until the end of the sentence is reache no more rules can be found by semantic or syntactic indexing.

Since the rule selection process occurs mainly through semantic indexing, se failures do not occur frequently. However, when semantic indexing cannot find a execute, and the parser is forced to use syntactic indexing, then semantic failu sometimes occur, because the rules are not always able to anticipate the se implications of the slot-fillings they perform. For example, consider the fol example:

Mary left after the rain stopped and walked home.

MOPTRANS's Conjunction Rule is used to conjoin "walked" and "walked," an assign "Mary" as the ACTOR of the PTRANS (The details of the Conjunction R be discussed in chapter 7). However, this rule encounters a semantic failure conjoining the right verbs.

The Conjunction Rule is indexed syntactically in this example, since the connection, between "Mary" and "walked," cannot be found during the sea semantic connections, due to the removal of "Mary" from active memory by the Rule. By syntactic indexing, the conjunction rule could match the syntactic pat active memory in two possible ways: "walked" could be conjoined to either "l "stopped." Since the rule is indexed syntactically in this situation, it does not know is the correct choice. The priority assignment of the Conjunction Rule causes the choice to be selected, since the rule prefers to conjoin things which are closer tog the sentence. Thus, the Conjunction Rule first attempts to conjoin "walked "stopped." This causes the attempted assignment of "rain" as the ACTOR of the WALK. However, this violates the prototype for what should be the ACTOR of Thus, a semantic failure occurs, and the parser returns to the state it was in bef execution of this rule was attempted. The execution of the Conjunction F "stopped" and "walked" is prohibited from occurring, and again the parser choose to be executed. This time, the Conjunction Rule is chosen again, but to conjoi and "walked." No semantic failure occurs upon execution of the Conjunction F second time, and "Mary" is assigned to be the ACTOR of WALK.

Semantic Failures and Prototype Failure Rules

Semantic failures also sometimes occur with rules which are normally indexed semantically. This is because some syntactic constructions do not always uniquely point to one slot-filling that should take place. For instance, in English we can say, "John killed Mary," "The explosion killed Mary," "The gun killed Mary," etc. In situations like these, it is not always possible for the MOPTRANS parser to find the correct rule to be executed through semantic indexing.

The representation that the MOPTRANS parser builds for "killed" is a state change called DIE. DIE consists of a change of one's HEALTH, which is measured on a numeric scale from +10 to -10. Thus, DIE is the state change of HEALTH from 0 (its default value) to -10.

Normally, the parser expects the CAUSE of the concept DIE to appear as its subject. Thus, "The shooting killed Mary" is parsed as (SHOOT CAUSE DIE). When the CAUSE of the death appears as the subject of the word "killed," then the Subject Rule is executed smoothly, and the correct representation is built. However, problems occur when the ACTOR of the CAUSE appears as the subject, as in "John killed Mary," or the WEAPON of the CAUSE, as in "The gun killed Mary." In these situations, semantic failure occurs, because "John" and "gun" do not meet the prototype for what the CAUSE of the concept DIE should be.

To handle cases like this, the MOPTRANS parser has special *Prototype Failure Rules*. These rules are indexed by certain slots of structures. For example, in the case of "John killed Mary," the Prototype Failure Rule is indexed under CAUSE and DIE, since the parser was trying to assign "John" as the CAUSE of DIE. Whenever the parser attempts to fill the CAUSE slot of DIE, and this slot-filling fails, then it indexes to see if it has any Prototype Failure Rules which cover this situation. In this case, there would is a Prototype Failure Rule, which is the following:

```
Prototype Failure Rule 1

Situation: Trying to fill the CAUSE slot of DIE
Failure:   Expected an ACTION, but the filler is a PERSON
Remedy:    Build an ACTION conceptualization, fill the ACTOR slot
              of the conceptualization with the PERSON
```

Using this rule, the MOPTRANS parser builds the representation (ACTION ACTOR JOHN CAUSE (DIE OBJECT MARY)) for the sentence "John killed Mary."

A similar Prototype Failure Rule exists for the situation encountered in parsing "The gun killed Mary":

```
Prototype Failure Rule 2

Situation: Trying to fill the CAUSE slot of DIE
Failure:   Expected an ACTION, but the filler is a WEAPON
Remedy:    Build an ACTION conceptualization that is associated
              with the weapon; fill the INSTRUMENT slot of
              this conceptualization with the WEAPON
```

One of the pieces of information that MOPTRANS has about GUNs is that they are often INSTRUMENTS of the action SHOOT. Thus, the parser builds the representation (SHOOT INSTRUMENT GUN RESULT (DIE OBJECT MARY)) for the sentence "The gun killed Mary," using the above Prototype Failure Rule.

Prototype Failure Rules are sometimes used, in conjunction with the concept refinement rules discussed in chapter 5, to resolve semantic ambiguities. This is because different senses of an ambiguous word sometimes require different slot-fillings. For example, consider the following sentences:

John left the restaurant.
John left a tip for the waitress.

"Leave" is ambiguous in these two examples in that it refers to a PTRANS in the first example, but to an ATRANS in the second example. To resolve this ambiguity, the dictionary definition of this word has pointers to two different nodes in the hierarchy: ATRANS and PTRANS. However, depending on which frame "leave" refers to, its syntactic OBJECT fills a different semantic slot. In the case of PTRANS, the direct object of "leave" fills the FROM slot. However, the object of "leave" meaning ATRANS is the semantic OBJECT of the ATRANS.

To handle this different slot assignment, Prototype Failure Rules are used. The dummy structure for "leave" is given a prototype for its OBJECT slot, which indicates that the OBJECT must not be a LOCATION. Given this prototype, when the syntactic Object Rule attempts to fill the OBJECT of "leave" with "restaurant," a semantic failure occurs (because all BUILDINGs are defined as LOCATIONs). However, a Prototype Failure rule exists, which is the following:

```
"Leave" Failure Rule

Situation: Trying to fill the OBJECT slot of "leave"
Failure:   Expected a movable PHYS-OBJECT, but the filler is a LOCATION
Remedy:    Replace the dummy representation of "leave" with
           the representation PTRANS.  Fill the FROM slot of
           PTRANS with the LOCATION.
```

For "leave" meaning ATRANS, the frame selection process proceeds as it did with "fix." The OBJECT of "leave" is filled in, and the concept refinement demons change the representation to ATRANS.

6.12 Generalized Syntactic Rules and Syntactic Ambiguities

The resolution of syntactic ambiguities is accomplished in the MOPTRANS parser through the Generalized Syntactic Rule indexing process. A syntactically ambiguous word has pointers in its dictionary definition to all of the syntactic classes to which it could belong. When the semantic or syntactic rule-indexing methods find a rule to execute which requires that the word in question belong to one of its possible syntactic classes, the ambiguity is resolved. The word is assigned to belong to the chosen syntactic category, and the rule is executed. For example, verbs which can either be past active or past participle, such as "called," are given pointers to the syntactic categories V and VPP. Then, depending on what Generalized Syntactic Rule is chosen by the indexing method, "called" is assigned to be one or the other class. If the Subject Rule is chosen to be executed, then "called" is automatically assigned to be a V. However, if semantics prefers the Unmarked Passive Rule, then "called" becomes a VPP.

This process also resolves ambiguities for words which are both syntactically and

semantically ambiguous. Pointers are placed in the dictionary definition of the ambiguous word to its possible semantic meaning, and to its possible syntactic classes. Pairs of pointers are linked, so that each syntactic pointer is paired with a semantic pointer, and vice versa. Then, when the ambiguous word is resolved syntactically, as described above through the rule selection process, the pairing of pointers causes the parser to also disambiguate the word semantically.

The Spanish word "armada" is disambiguated in this way in the MOPTRANS parser. As a noun, "armada" means army, but it is also the past participle form of the verb "to arm," and thus can be used as the English "armed with ...". In the definition of "armada" are pointers to the concepts ARMY and ARMED-WITH, and the syntactic categories N and VPP. The pointers are linked, ARMY with N and ARMED-WITH with VPP. The disambiguation of "armada" then takes place when a Generalized Syntactic Rule is selected which required that "armada" be a member of only one of its possible syntactic categories. Thus, in the context "La batalla peleada por la armada ..." (the battle fought by the army ...), the Spanish Noun Phrase Rule would match on the pattern of a determiner followed by a noun, and "armada" would be assigned as a N. At the same time, the pairing of pointers would cause the parser to build the conceptual representation ARMY. On the other hand, in a context like "la patrulla armada con una pistola ..." (the patrolman armed with a pistol...), the only Generalized Syntactic Rule that would apply would be the Spanish Unmarked Passive Rule, and "armada" would be assigned to be a VPP, thus causing the conceptual representation ARMED-WITH to be built.

6.13 Conclusion

Request-based syntactic knowledge used in many previous integrated parsers suffered from two problems. First, this knowledge was mixed in with other types of knowledge, such as semantic knowledge and disambiguation knowledge. Although this allowed for syntactic and semantic processing to be easily integrated in these parsers, it also resulted in an inefficient representation of syntactic knowledge. Second, the local syntax-checking performed by lexically-based requests was not adequate for complex syntactic constructions. These constructions required numerous requests, and even then examples could be found for which these requests would not work.

In this chapter, I have presented an approach to syntactic knowledge that does not suffer from these problems. First, since MOPTRANS stores this knowledge in terms of general rules, the inefficiencies in representation suffered by lexically-based syntactic rules are avoided. This results in the need for fewer rules, and also allows knowledge to be shared among languages. Although the knowledge base in this approach to syntax is not integrated, syntactic and semantic processing are still completely integrated, because of the semantic way in which Generalized Syntactic Rules are indexed. Thus, this approach has the advantages of integrated processing, without the inefficiencies of a totally integrated knowledge base.

Second, this approach allows for the necessary syntactic representations to be built to handle more complex English constructions. The principle of conceptual analysis that syntactic analysis should only be performed in service of semantic analysis still holds, but the parser categorizes constituents according to their syntactic function, such as main verbs, relative clause verbs, etc., because this enables the processing of syntactically complex sentences such as "The soldier called to the sergeant shot in the arm was

reprimanded," with simpler rules than are needed with lexically-based syntactic knowledge.

7. Multiple Language Parsing

In a multi-lingual parser, one of the major benefits of the separation of syntactic and semantic knowledge is the ability to share knowledge across languages. In this chapter, I will discuss the knowledge which the MOPTRANS parser is able to share between some or all of its input languages.

7.1 Conceptual Knowledge

MOPTRANS' conceptual knowledge base is used, completely unchanged, to parse all five of the input languages. This tends to confirm the assertion that this knowledge is language-independent, or largely non-linguistic in nature. In addition, the concept refinement demons which I discussed in chapter 5 are also identical for the parsing of each of the languages.

We have already seen examples of the operation of the concept refinement demons in conjunction with the conceptual hierarchy for Spanish and English. The same rules operate in the parsing of French, German, and Chinese, to refine vague or ambiguous representations as more information is provided during the parsing process. Let us look at an example of the concept refinement rules in action in each of these languages. First, consider the following French story, and the representation and translation produced by MOPTRANS:

French: Les ambulances de la Croix Rouge ont transporte d'urgence deux jeunes filles, dont les mains avaient ete blessees par suite d'une bombe, a l'hopital Manolo Morales.

```
Final representation:

EXPO =
 CONCEPT EXPLODE-BOMB
 INST    OBJ6 =
           CONCEPT BOMB
           INST-OF EXPO
 OBJECT  HUM21 =
           CONCEPT PERSON
           GENDER  FEMALE
           B-PART  OBJ5 =
                     CONCEPT BODYPART
           AGE     YOUNG
           NUMBER  2
 RESULT  INJ2 =
           CONCEPT   INJURED
           R1        HUM21
           RESULT-OF EXPO
```

```
PTR99 =
 CONCEPT PTRANS-BY-AMBULANCE
 OBJECT  HUM21
 TO      LOC7 =
           CONCEPT HOSPITAL
 INST    OBJ4 =
           CONCEPT  AMBULANCE
           OWNED-BY ORG5 =
                      CONCEPT MEDICAL-ORG
                      OWNS    OBJ4
                      #NAME   RED CROSS
           INST-OF  PTR99
 PRECOND INJ2

Total time:  80114 msecs.
NIL
```

English translation: 2 young women who were injured by a bomb in the hands were rushed by an ambulance owned by the Red Cross to the hospital.

German translation: 2 junge Frauen wurden nach das Spital mit einem Krankenwagenen von dem Rotkreutz, gehastet. Sie wurden mit einer Bombe verwundet.

In this story, the French "ont transporte d'urgence" (have transported urgently) first causes the parser to build the representation PTRANS (the Conceptual Dependency primitive for change in physical location). "Transporter" expects its subject to fill the ACTOR slot of the PTRANS, so when the French Subject Rule assigns "les ambulances" as the subject of "transporter," a Prototype Failure Rule comes into play. This failure rule causes the structure AMBULANCE to be placed into the INSTRUMENT slot of the PTRANS, since AMBULANCE is a VEHICLE. At this point, the Slot-filler Specialization demon refines the structure PTRANS, because the filling of the INSTRUMENT slot with AMBULANCE matches the expected prototype for the INSTRUMENT slot of a more specific structure, called PTRANS-BY-AMBULANCE. This new structure is part of an event sequence, called M-HOSPITAL, which is the following:

M-HOSPITAL = INJURY + PTRANS-BY-AMBULANCE + TREATMENT

This event sequence is activated. Finally, when the parser reads "blesees" (injured), this matches the first event of M-HOSPITAL, which is the PRECONDITION of the PTRANS-BY-AMBULANCE. Thus, the Expected Event Specialization Demon causes the PRECONDITION relation to be assigned between the structure INJURED and PTRANS-BY-AMBULANCE.

This same sort of concept refinement also occurs during the parsing of Chinese stories. Consider the following example, which is parsed by MOPTRANS:

Chinese: yilang jintian shuo , yilake tewu xiji yilake bianjing, dasi le er ren , zhuazou le xuduo renzhi.

Literal English: Iran today say, iraqi agents attack iraqi border, kill (past marker) 2 men, seize (past marker) a number of hostages.

Good English: Iran today said iraqi agents killed two men and seized a number of hostages in an attack on the iraqi border.

```
Final representation:

HAR9 =
 CONCEPT HARM
 OBJECT  LOC6 =
           CONCEPT LOCATION
           TYPE    BORDER
 ACTOR   HUM16 =
           CONCEPT      PERSON
           TYPE         AGENT
           NATIONALITY  LOC4=
                          CONCEPT NATION
                          #NAME   (IRAQ)
GET2 =
 CONCEPT TAKE-HOSTAGES
 ACTOR   HUM16
 OBJECT  HUM18 =
           CONCEPT HOSTAGE
           NUMBER  A-NUMBER-OF
HAR10 =
 CONCEPT HARM-PERSON
 ACTOR   HUM16
 OBJECT  HUM2 =
           CONCEPT PERSON
           NUMBER  2
 RESULT  DEA0 =
           CONCEPT DEAD
           R1      HUM2
MTR1 =
 CONCEPT MTRANS
 ACTOR   HUM15 =
           CONCEPT   PERSON
           SPOKESMAN LOC5 =
                       CONCEPT NATION
                       #NAME   (IRAN)
 TIME    INS1 =
           CONCEPT INSTANCE
           DAY     TODAY
 OBJECT  HAR9


Total time:  47299 msecs.
NIL
```

English translation: Iran said today that Iraqi agents killed 2 men. The agents seized a number of hostages during a raid near the border with Iraq.

German translation: Iran sagte heute dass ein irakischer Agent 2 Maenner

toeteten. Sie nahm mehrere Geisel.

The event TAKE-HOSTAGES resulted from the Chinese text "zhuazou le xuduo renzhi," because of the concept refinement process. "Zhuazou" (seize) is defined as a GET-CONTROL. When the OBJECT of the GET-CONTROL is filled with the concept HOSTAGE (from the Chinese word "renzhi") by the Chinese Verb Phrase Rule (which will be discussed later in this chapter), the Slot-filler Specialization Demon changes the representation GET-CONTROL to TAKE-HOSTAGES.

Concept refinement rules perform this sort of inferencing in the parsing of German, also. In addition, these rules are used to anticipate verbs. In German, verbs appear at the end of subclauses, or in any clause which uses auxiliaries. Here are some examples:

German: John sagte dass Mary mit dem Zug nach Frankfurt reiste.

English: John said that Mary traveled by train to Frankfurt.

German: Eine Prisoner wurde von einer Executionstruppe zurechtgestellt.

English: A prisoner was executed by a firing squad.

In the first example, since the verb "reiste" (travelled) occurs in the clause beginning with "dass" (that), it appears at the end of this clause, after its subject, its object (although this particular verb has no object), and any prepositional phrases. This is the case with the past participle in the second example, "zurechtgestellt" (executed), since it is used here with the passive auxiliary, "wurde." The auxiliary appears in the position normally occupied by the verb, and the participle is moved to the end of the clause.

In cases like these, the action or state which the subclause describes is not explicitly mentioned until the end of the clause. However, often it is the case that the action to which the clause refers can be inferred before the reader actually gets to the verb. This is because the slot-fillers specified by the noun groups and prepositional phrases appearing before the verb, and the semantic roles which they play, can provide enough information to infer the action in which they are playing a role. This is the case in the two examples above. In the first example, since the reader has specific knowledge about the action with which a train is normally associated (a PTRANS), and since the prepositional phrase "nach Frankfurt" (to Frankfurt) can fill the DESTINATION slot in a PTRANS, the reader can infer that the verb at the end of the clause will mean PTRANS. Likewise, since a firing squad ("Executionstruppe") often executes prisoners, the reader can anticipate in the second example that the past participle will refer to some sort of killing, before the verb is actually read.

The MOPTRANS parser can anticipate verbs in German, using the same concept refinement demons which are used in the parsing of all its languages. In order to anticipate the verb of a clause, the MOPTRANS parser builds a "dummy" action whenever it encounters the beginning of a clause. Then, German noun group attachment rules (which will be discussed later in this chapter) are used to attach noun groups or prepositional phrases to the dummy action. As more information is added to the dummy action, the refinement rules can replace the vague dummy representation with particular actions, if the slot-fillings provide enough information. This occurs in both of the above examples. In the first example, a dummy representation, ACTION, is built when the parser encounters "dass" (that). Then, "mit dem Zug" (with the train) is attached to this dummy representation. This causes the representation TRAIN to fill the INSTRUMENT slot of ACTION, since the preposition "mit" (with) refers to this slot. Because of this slot-

filling, the Slot-filler Specialization demon changes the dummy representation to PTRANS, since the expected slot-filler of the INSTRUMENT slot of PTRANS is a VEHICLE, and TRAIN is a VEHICLE. The prepositional phrase "nach Frankfurt" (to Frankfurt) is also attached to the PTRANS structure, causing the TO slot of the PTRANS to be filled with (CITY NAME FRANKFURT). Thus, when the verb is finally encountered, the parser has already built the representation (PTRANS ACTOR MARY INSTRUMENT TRAIN TO (CITY NAME FRANKFURT)).

In the second example above, the dummy representation ACTION is built when the passive auxiliary "wurde" is encountered. The parser knows that a FIRING-SQUAD is an ORGANIZATION whose MEMBERS are likely to perform the action EXECUTE. Thus, when the prepositional phrase "von einer Executionstruppe" (by a firing squad) is read, and FIRING-SQUAD is attached to the dummy representation, the Slot-filler Specialization demon causes the dummy action to be refined to EXECUTE. Thus, again the action expressed by the verb is inferred by the parser before the verb is encountered.

Prototype Failure Rules

Many of the MOPTRANS parser's Prototype Failure Rules are used in the parsing of all of the input languages. I presented examples of Prototype Failure Rules in chapter 6, along with English sentences in which these rules applied. The same rules are applicable to the system's other languages. For example, two rules which I presented in chapter 6 were the following:

```
Prototype Failure Rule 1 (for "John killed Mary")

Situation: Trying to fill the CAUSE slot of DIE
Failure:   Expected an ACTION, but the filler is a PERSON
Remedy:    Build an ACTION conceptualization, fill the ACTOR slot
              of the conceptualization with the PERSON


Prototype Failure Rule 2 (for "The bullet killed Mary")

Situation: Trying to fill the CAUSE slot of DIE
Failure:   Expected an ACTION, but the filler is a WEAPON
Remedy:    Build an ACTION conceptualization that is associated
              with the weapon; fill the INSTRUMENT slot of
              this conceptualization with the WEAPON
```

These rules are needed because of the ability in English to use several possible slots as the subject of "killed." The subject can be the CAUSE of the death, the ACTOR of the CAUSE, or the WEAPON used in the CAUSE. Thus, the regular Subject Rule does not suffice for "killed," because the same slot is not filled by the subject of the verb.

These failure rules are used in the other languages in the system, also, as the following examples illustrate:

French:
Jean a tue Marie. (John killed Mary.)
Marie a ete tuee par un coup de feu. (Mary was killed by a shot.)
Marie a ete tuee par une explosion. (Mary was killed by an explosion.)

Spanish:
Juan mato a Maria. (John killed Mary.)
La bala mato a Maria. (The bullet killed Mary.)
La explosion mato a Maria. (The explosion killed Mary.)

German:
John toetete Mary. (John killed Mary.)
Der Kugel toetete Mary. (The bullet killed Mary.)
Die Explosion toetete Mary. The explosion killed Mary.)

Chinese:
JangSan shale MeiLi. (John shot-dead Mary.)
MeiLi bei zedan sha-se. (Mary was shot-dead by the bullet.)
MeiLi bei bauja shuai-se. (Mary was crushed-dead by the explosion.)

We see that similar constructions are possible in French, Spanish, German, and Chinese. Thus, the Prototype Failure Rules for these examples are the same as for English.

Another Prototype Failure Rule was presented earlier in this chapter, which was responsible for filling the INSTRUMENT slot of a PTRANS verb with a VEHICLE when the VEHICLE appears as the subject of the verb. This rule also applies to the other languages, which permit the same construction:

English: Red Cross ambulances rushed two young women whose hands had been injured as the result of a bomb to Manolo Morales hospital.
Spanish: Ambulancias de la Cruz Roja trasladaron al hospital Manolo Morales a dos jovencitas que sufrieron mutilaciones de sus manos a causa de explosion de una bomba.
French: Les ambulances de la Croix Rouge ont transporte d'urgence deux jeunes filles, dont les mains avaient ete blessees par suite d'une bombe, a l'hopital Manolo Morales.
German: Ein Rotkreutzkrankenwagen hastete 2 junge Frauen deren Haende von einer Bombe verwundet wurden nach Manolo Morales Spital.
Chinese: hongshizi jijiuche jiang zai yi ci baozha shijian zhong zhashang le shou de er ming nianqing de funu jisu song wang mannuoluo molaersi yiyuan.

Since in each language, the subject of the verb meaning PTRANS is AMBULANCE, the same Prototype Failure Rule applies to each language.

Here is another example of a Prototype Failure Rule:

```
Nation-actor Prototype Failure Rule:

Situation: Trying to fill the ACTOR slot of an MTRANS
Failure:   Expected a PERSON, but the filler is a NATION
Remedy:    Build a PERSON structure, fill the ACTOR slot
              of the conceptualization with the PERSON, assign
              the PERSON to be a SPOKESMAN of the NATION
```

This rule is for sentences like "Iran said today that it had destroyed three oil tankers." The structure MTRANS, the Conceptual Dependency primitive for transfer of information, expects a PERSON as its ACTOR. However, in this sentence, a NATION,

"Iran" is syntactically in the position normally occupied by the ACTOR of the MTRANS. Thus, the Prototype Failure Rule above is needed.

Again, it is allowable in all of the languages which MOPTRANS can parse to use a nation as the ACTOR of an MTRANS, to mean that a spokesman for the nation conveyed some information. Thus, this Prototype Failure Rule is applicable to all of MOPTRANS' input languages.

Prototype Failure Rules express linguistic knowledge, not conceptual knowledge. For instance, the fact that we can express (DEAD OBJECT MARY CAUSE (SHOOT ACTOR JOHN OBJECT MARY)) by saying "John killed Mary by shooting her," or "The shot fired by John killed Mary," is a fact about language, not a conceptual fact. Given that this knowledge is linguistic, one would expect the specifics of this knowledge to vary from language to language. However, this is not the case. Most of the Prototype Failure Rules used in MOPTRANS are applicable to all of the languages in the system. One possible explanation for this is that although Prototype Failure Rules express linguistic knowledge, this knowledge involves very basic linguistic inferences. One rather basic rule of language in general seems to be that it is not necessary to explicitly say something that is easily inferable. Prototype Failure Rules appear to be particular instances of this general rule. For example, in "John killed Mary," there are very few roles which "John" could be playing in the action expressed by "killed." Thus, it seems acceptable in all of the languages in the MOPTRANS system to say "John killed Mary" and expect the reader to be able to infer that this means (DEAD OBJECT MARY CAUSE (ACTION ACTOR JOHN)). Similarly, although nations are not animate objects and thus cannot be the ACTORs of actions, it is acceptable to use a nation as the subject of some verbs in many languages, because it is easy to infer that it must be some person playing a particular role for the nation who actually performed the action. Since these inferences seem quite basic and far-removed from the particular language being used, this may explain why Prototype Failure Rules do not vary more from language to language.

7.2 Shared Syntactic Knowledge in MOPTRANS

Although the syntactic knowledge expressed in Generalized Syntactic Rules is mainly linguistic in nature, and would thus be expected to vary from language to language, many of these rules are shared between two or more of the languages in MOPTRANS. This sharing of linguistic knowledge reflects the similarities between the languages in the system. In this section, I will discuss the rules which the parser shares between two or more languages.

In all, the MOPTRANS parser uses approximately 285 Generalized Syntactic Rules to parse English, Spanish, French, German, and Chinese. Figure 7-1 shows how many of these rules are shared between languages. In all, about 44% of MOPTRANS' parsing rules apply to more than one language. Very few rules are used in all languages, due to the fact that Chinese is so distinct from all of the other languages in the system. On the other hand, a large percentage of the shared rules are shared by at least 3 languages, reflecting the similarities between English, Spanish, and French. German also shares a number of parsing rules with these languages, although its freer word order requires different sets of rules for parsing clauses, sentences with auxiliaries, and various other constructions.

```
Total number of Generalized
Syntactic Rules in MOPTRANS:  285

Number of languages          Number of rules
rules are applicable to

          1                        161
          2                         42
          3                         54
          4                         24
          5                          4
```

Figure 7-1: Rules Shared Between Languages in the MOPTRANS Parser

7.2.1 Generalized Syntactic Rules Which Apply to All Languages

7.2.1.1 Conjunction

Conjunction is a construction which is highly ambiguous in any language. Almost anything can be conjoined, as long as it is syntactically similar and fills the same semantic role in the sentence. A conjunction can join two noun phrases, or two verb phrases, or two sentences, or even two verbs phrases with adverbs in front of them.

To deal with these ambiguities, the MOPTRANS parser uses the same conjunction rules for the five languages which it can parse. To illustrate these rules, I will first discuss English examples of various different syntactic constituents conjoined together, and discuss how these examples are handled by the system's conjunction rule. Here are the examples:

I saw John and Mary.
In the park I saw John and Mary saw Bill.
I saw John and heard Mary.
I slowly walked to the store and quickly ran back.

The conjunction rule is as follows:

```
Conjunction Rule

Syntactic pattern:         CONST1, "and", CONST2
Additional restrictions:   CONST1 and CONT2 are of the same syntactic class,
                              and the n (n>= 1) most recent Generalized
                              Syntactic Rules applied to CONST1, for which
                              CONST1 was the rightmost element operated on
                              by the rules, can apply to CONST2
Syntactic assignment:      CONST1 CONJOINS CONST2
Semantic action:           Apply the Generalized Syntactic Rules specified
                           in the additional restrictions to CONST2
Result:                    CONST2
```

To show how this rule works, consider the first example above, "I saw John and Mary." The following rules would have been executed on the word "John": the Noun Phrase rule, which recognizes a noun standing alone as a noun phrase; and the Direct Object rule, which assigns "John" to be the object of "saw." The Direct Object rule

would be the most recently executed rule. Thus, when the MOPTRANS parser parses "Mary," it tries to find something before "and" whose rules can be executed on "Mary." First, it finds nothing, since "Mary" is just a noun. However, after the Noun Phrase rule is executed on "Mary," then "John" matches as a constituent which can be conjoined with "Mary," since the most recently executed rule, the Direct Object rule, can be executed on "Mary" also. Thus, the representation produced is the same as would be produced by "I saw John. I saw Mary."

In the second example above, "In the park I saw John and Mary saw Bill," the same thing happens at first, since "Mary" can match with "John." Thus, a backup rule is needed, which is the following:

```
Conjunction Backup Rule

Syntactic pattern:          NP, V
Additional restrictions:    NP is CONJOINED with another NP
Action:                     Back up to the execution of the Conjunction Rule
```

Thus, in the parsing of "In the park I saw John and Mary saw Bill," the parser first incorrectly conjoins "John" and "Mary." Then, when the word "saw" is found, the parser backs up, and assigns "Mary" as the ACTOR of "saw." After this backup, the parser tries to conjoin the two instances of "saw." The most recent rule applied to the first instance of "saw" is the Direct Object rule, but since "John," and not "saw," was the rightmost element operated on by this rule, this rule does not qualify as one to also be applied to the second instance of "saw." The most recent rule executed on the first "saw" which meets the qualifications of the conjunction rule is a prepositional phrase attachment rule (which will be discussed later in this chapter), which assigned the LOCATION of "saw" to be "park." Thus, this rule is executed on the second instance of "saw," also, and the two sentences are conjoined.

In the third example above, "I saw John and heard Mary," the matching rules identify "saw" and "heard" as of the same syntactic class. Again, the Direct Object rule is the most recent rule run on "saw," but it does not apply, since "saw" is not the leftmost element in this rule. Instead, the Subject rule, which was run on "saw" to assign "I" as the ACTOR of "saw," is the rule which applies to "heard." The two verbs are conjoined, and "I" is assigned to be the ACTOR of "heard," also.

Finally, in the fourth example, "I slowly walked to the store and quickly ran back," the Adverb rule first operates on "quickly" and "ran," attaching the property SPEED FAST to the conceptualization built by "ran." When the conjunction rule tries to match "ran" and "walked," the most recent rule for which "walked" was the rightmost element is again the Subject rule. Thus, "I" is assigned to be the ACTOR of "ran." Then, the Adverb rule, which was executed on "slowly" and "walked," is run on "ran," also. However, this rule fails, because "ran" already has the property SPEED FAST, which contradicts the word "slowly." In a sentence like "I slowly walked to the store and looked at the merchandise," the adverb rule would succeed, and the property SPEED SLOW would be added to the representation of "looked," also.

The Conjunction Rule is not infallible. For example, consider the following sentence:

I know John and Mary saw Fred this morning.

The preferable parse for this sentence is the same as "I know that John and Mary saw Fred." However, this sentence would be parsed by MOPTRANS as two sentences conjoined together. This is because MOPTRANS does not have an idea as to what sorts

of conceptual entities can be conjoined. It seems that coinjoining "know" and "sa rather awkward; therefore, people prefer the alternate interpretation of this sent However, MOPTRANS does not have the knowledge necessary to make this so judgement. I will have more to say about this in chapter 8.

This same Conjunction Rule is used for parsing conjunctions in French, Spa German, and Chinese. Here are some examples from other languages:

German: Iran sagte heute dass irakische Agenten waehrend eines Angriffes in der Naehe von der irakischen Grenze 2 Maenner toeteten und mehrere Geisel nahmen.

English: Iran today said iraqi agents killed two men and seized a number of hostages in a raid near the border with iraq.

In this sentence, the Conjunction Rule matches on "toeteten" (killed) and "nah (took). During the parse of the sentence before the conjunction, the German attachment rules (which will be described later in this chapter) assign "Agenten" (ag to be the ACTOR of "toeteten" (killed), and "2 Maenner" (2 men) to be the OBJ Also, a rule is executed which assigns the clause beginning with "dass" as the OBJEC the MTRANS. When the Conjunction Rule is executed, the parser attempts to dupl these three assignments with "sagte." It succeeds in assigning "Agenten" as the ACT of "nahmen," and assigning "nahmen" as the OBJECT of the MTRANS. How before the Conjunction Rule, an NP attachment rule is run on "nahmen" and "G (hostages), assigning HOSTAGE as the OBJECT of the GET-CONTROL (the assign of "Geisel" as the OBJECT of "nahmen" will be discussed later in the chapter). since the OBJECT slot of GET-CONTROL is already full, the Conjunction Rule fa fill the OBJECT slot of the GET-CONTROL with "2 Maenner." This results in following final representation of the sentence:

```
Final representation:

GET0 =
 CONCEPT TAKE-HOSTAGES
 OBJECT  HUM9 =
           CONCEPT HOSTAGE
           NUMBER  SEVERAL
 ACTOR   HUM7 =
           CONCEPT     PERSON
           NATIONALITY LOC7 =
                         CONCEPT NATION
                         #NAME   (iraq)
HAR3 =
 CONCEPT HARM-PERSON
 ACTOR   HUM7
 OBJECT  HUM8 =
           CONCEPT PERSON
           GENDER  MALE
           NUMBER  2
 RESULT  DEA1 =
           CONCEPT   DEAD
```

```
          R1          HUM8
          RESULT-OF HAR3
          DURING    HAR2 =
                      CONCEPT     HARM
                      PLACE       LOC8 =
                                    CONCEPT LOCATION
                                    NEAR    LOC10 =
                                              CONCEPT    LOCATION
                                              NATION-ADJ LOC9 =
                                                           CONCEPT NATION
                                                           #NAME
(iraq)                SETTING-FOR DEA1
MTRO =
 CONCEPT MTRANS
 ACTOR   HUM6 =
          CONCEPT   PERSON
          SPOKESMAN LOC6 =
                      CONCEPT NATION
 TIME    INS3 =
          CONCEPT INSTANCE
          DAY     TODAY
 OBJECT  DEA1

Total time:  82692 msecs.
NIL
```

Conjunction in Chinese is more limited than in MOPTRANS' other languages. "Gen," the Chinese word corresponding to "and," is only used to conjoin noun groups. Verb phrases are not usually conjoined, but rather are simply strung together without any explicit conjunction markers, as in the example discussed earlier:

Chinese: yilang jintian shuo , yilake tewu xiji yilake bianjing, dasi le er ren , zhuazou le xuduo renzhi.

Literal English: Iran today say, iraqi agents attack iraqi border, kill (past marker) 2 men, seize (past marker) a number of hostages.

Good English: Iran today said iraqi agents killed two men and seized a number of hostages in an attack on the iraqi border.

Thus, the Conjunction Rule used in the other languages applies to noun phrase conjunctions only. The processing of strings of verb phrases will be discussed later in this chapter.

7.2.1.2 Pronominal Reference

Strategies for resolving pronominal reference are shared among the languages in the system, also[1]. These consist of the following two strategies: resolution by semantic

[1] The MOPTRANS parser has not been run on Chinese examples containing pronouns.

inference, and resolution by syntactic role.

Resolution by Semantic Inference

Charniak argued in (Charniak, 1972) that pronominal reference should *often* be performed in the course of normal semantic processing that must go on in natural language understanding anyway. This is often the case in the MOPTRANS parser. Consider the following example:

> A policeman was shot and critically wounded by a terrorist. Ambulances rushed him to the hospital, where he underwent emergency surgery.

In this example, the referent of the pronoun "him" is determined by the role-binding information in the event sequence HOSPITAL, which is the following:

```
HOSPITAL = INJURY + PTRANS-BY-AMBULANCE + TREATMENT
```

The word "rushed" is defined as a PTRANS. Then, the Expected Event Specialization Demon matches "rushed" with the expected event PTRANS-BY-AMBULANCE, since the wounding mentioned in the first sentence causes HOSPITAL to be activated, and since PTRANS-BY-AMBULANCE is a type of PTRANS. Once "rushed" has been determined to refer to the event PTRANS-BY-AMBULANCE, HOSPITAL provides the role-binding information that the OBJECT of the PTRANS-BY-AMBULANCE is the same at the OBJECT of the INJURY. Thus, at this point the parser infers that the policeman is the semantic OBJECT of "rushed." When the parser reads "*him*," *the reference* is already resolved.

Resolution by Syntactic Role

The other half of MOPTRANS' pronominal reference resolution is a set of pronominal reference checks which reside in the various rules that attach noun phrases to verbs or prepositions. These checks vary somewhat, depending on the particular syntactic role that the pronoun plays in the sentence.

To explain this strategy, consider the following example:

> The soldier killed the man who shot him in the arm.

Since "him" is not a reflexive pronoun, and since "the man" is the subject of "shot" and "him" is the direct object of this verb, we know that "him" cannot refer to "the man." Therefore, it must refer to "the soldier," since "the soldier" is the only other possible referent[2].

To handle this sentence, the following check is included in the Object Rule:

---

[2]*This is not always true.* "Him" might refer to some third person, mentioned earlier in the context, as in "The soldier's friend was bleeding to death. Out of revenge, the soldier killed the man who shot him." However, as we will see in the next section, the MOPTRANS parser is sensitive to contextual information from other sentences. Thus, in this case, the parser would not choose "the man" as the referent of "him."

Object Rule

```
Syntactic pattern:         S, NP
Additional restrictions:   NP is not attached syntactically
Syntactic assignment:      NP is (syntactic) OBJECT of S
Semantic action:           NP is (semantic) OBJECT of S (or another
                              slot, if specified by S)
                           If NP is a non-reflexive PRONOUN, find another
                              NP from earlier in the story with the
                              appropriate semantic restrictions
                              which is not the SUBJECT of S.
                              If only one of these exists, change
                              the representation of NP to this
                              NP's representation.
Result:                    S, NP
```

Similar checks are included as part of the action of the Subject Rule, and the Prepositional Phrase attachment rules in English:

Subject Rule

```
Syntactic pattern:         NP, V (active)
Additional restrictions:   NP is not already attached syntactically
Syntactic assignment:      NP is SUBJECT of V, V is a MAIN CLAUSE
Semantic action:           NP is ACTOR of V (or another slot, if specified
                              by V)
                           If NP is a PRONOUN, find another NP from earlier
                              in the story with the appropriate semantic
                              restrictions. If only one of these exists,
                              change the representation of NP to this
                              NP's representation.
Result:                    V (changed to S)
```

Prepositional Phrase Attachment Rule 1

```
Syntactic pattern:         NP or S, PP
Additional restrictions:   none
Syntactic assignment:      PP is attached to NP or S
Semantic action:           Fill the slot (specified by the preposition)
                             of the NP or S with the NP in the PP
                           If NP is a non-reflexive PRONOUN, find another
                              NP from earlier in the story with the
                              appropriate semantic restrictions
                              which is not the SUBJECT of S.
                              If only one of these exists, change
                              the representation of NP to this
                              NP's representation.
Result:                    NP or S, NP in PP
```

Pronominal reference is handled in an identical way in German. Consider the

following example, which is parsed by MOPTRANS:

German: Ein Verbrecher wurde von dem Polizisten der ihn von Tierra Azul hierher fuhr, getoetet.

English: A criminal was killed by the patrolman who was driving him here from the city of Tierra Azul.

The referent of the pronoun "ihn" is found in much the same way as for English examples, upon execution of the German rule which attaches "ihn" to the verb "fuhr" (driving). The rule responsible for making this attachment is the following:

```
German NP Before V Rule

Syntactic pattern:        NP, V
Additional restrictions:  none
Syntactic assignment:     Assign the NP as the case-filler of the V,
                             according to the case of the NP
Semantic action:          If the V specifies that the case of the NP
                             should fill a particular slot, then fill
                             that slot with the NP.  Otherwise, perform
                             the default slot-filling associated with the
                             NP's case.
                          If NP is a non-reflexive PRONOUN, find another
                             NP from earlier in the story with the
                             appropriate semantic restrictions (if the
                             pronoun is accusative, then the second NP
                             cannot be the SUBJECT of the S).
                             If only one of these exists, change
                             the representation of NP to this
                             NP's representation.
Result:                   V
```

This rule will be discussed in detail in a later section of this chapter. The point of showing the rule now is to show that the pronominal reference portion of the rule is the same as for the English Object Rule.

Pronominal reference proceeds in much the same way in Spanish and French, also, although some modification is required due to the location of pronouns in these languages. Consider the Spanish version of the German example above:

Spanish: El reo Roger Fidel Morales Gonzalez fue matado por la patrulla que lo conducia en una camioneta desde Tierra Azul hacia esta ciudad.

English: A criminal, Roger Fidel Morales Gonzalez, was killed by the patrolman who was driving him in a car from Tierra Azul to this city.

In Spanish and French, object pronouns come before the verb. Thus, a special Pronoun rule is necessary. Here is the rule for Spanish:

Spanish Object Pronoun Rule

Syntactic pattern: OP (object pronoun), V
Additional restrictions: none
Syntactic assignment: none
Semantic action: none
Result: V, OP

This rule simply switches the order of the pronoun and the verb, so that normal object-attaching rules can operate to attach the pronoun to the verb. These object-attaching rules have checks, identical to the checks which are in the English rules. Thus, when the Spanish Direct Object Rule attaches "lo" (him) as the OBJECT of the PTRANS, the same restrictions apply to what can be the referent of the pronoun as applied for English. Since this pronoun is the object of the verb, and it is not reflexive, the verb's subject cannot be the referent. This leaves only one possible referent, the criminal ("el reo"). Thus, the referent of "lo" is resolved to be the criminal.

7.2.1.3 Other Reference Rules

Often, in multi-sentence stories, the same events or characters are mentioned more than once in the story. Thus, it is important for the parser to be able to identify when the same referent is referred to more than once.

To resolve multiple references to the same event or character, the MOPTRANS parser uses a conceptual memory, which is separate from its active memory. The conceptual memory contains all of the conceptual representations built during the story so far. This memory is referred to by resolution rules to check past events or characters to see if the current word or phrase might refer to any of them. The rules which MOPTRANS uses are the following:

Verb Referent Rule

Syntactic pattern: S
Additional restrictions: none
Syntactic assignment: none
Semantic action: Find any actions in conceptual memory from earlier in the story which can be merged with the S. If only one of these exists, merge the two representations.
Result: S

Definite NP Referent Rule

Syntactic pattern: NP
Additional restrictions: NP must have a definite pronoun, or be a proper noun
Syntactic assignment: none
Semantic action: Find any conceptualizations from earlier in the story which can be merged with the NP. If only one of these exists, merge the two representations.
Result: NP

Mergable conceptualizations are either conceptualizations which are the same, or conceptualizations which have IS-A relationships linking them. Thus, the Verb Referent Rule operates on the following two examples:

> Terrorists shot and killed a policeman. The policeman was gunned down when he tried to stop the terrorists from detonating a bomb.
>
> Police have arrested a terrorist responsible for the bombing of a Paris restaurant. The terrorist was captured after an intense search covering 3 square miles of the city.

In the first example, the Verb Referent Rule matches "gunned down" with "shot" because they both build the same representation, SHOOT. In the second example, this rule matches "captured" with "arrested" through IS-A links. "Captured" builds the representation GET-CONTROL, while "arrested" build ARREST. These two concepts are connected with an IS-A link from ARREST to GET-CONTROL. Thus, they match, and the Verb Referent Rule merges the two representations.

7.2.2 Rules Shared Between Similar Languages

Due to the similarities between English, Spanish, and French syntax, many of the rules which MOPTRANS uses for these languages are shared. In this section, I will discuss subject-verb-object construction, the use of prepositional phrases, and clause constructions for these three languages, and the shared rules which MOPTRANS uses for them. The rules for handling these syntactic constructions in German and Chinese will be discussed later in the chapter.

7.2.2.1 Subjects and Direct Objects

Identical subject and object rules are used in English, Spanish, and French. These rules were discussed in chapter 6. One additional rule is required for Spanish and French, because these languages sometimes allow for the subject to be placed after the verb. Here is an example:

> Spanish: Todavia se encuentra internada en el hospital la joven Rosa Areas, la que fue herida de bala por un uniformado.
>
> English: Rosa Areas is still in the hospital after being shot and wounded by a soldier.

In this sentence, the subject, "joven" (young person), is found after the verb, "se encuentra" (finds herself). To handle situations like this, the following rule is used for French and Spanish:

```
Inverted Subject Rule

Syntactic pattern:        V, NP
Additional restrictions:  V does not have a subject or a direct object
Syntactic assignment:     NP is SUBJECT of V
Semantic action:          NP is ACTOR of V (or another slot, if specified
                            by V)
Result:                   V (changed to S)
```

7.2.2.2 Prepositional Phrases

Identical rules are used to process prepositional phrases in English, Spanish, and French. They are the following:

```
Prepositional Phrase Rule

Syntactic pattern:        PREP, NP
Additional restrictions:  none
Syntactic assignment:     NP is PREP-OBJECT of PREP
Semantic action:          NP will be filler of the slot specified by the
                            PREP in some conceptualization
Result:                   NP or V, NP in PP

Prepositional Phrase Attachment Rule 1

Syntactic pattern:        NP or S, PP
Additional restrictions:  none
Syntactic assignment:     PP is attached to NP or S
Semantic action:          Fill the slot (specified by the preposition)
                            of the NP or S with the NP in the PP
Result:                   NP or S, NP in PP

Prepositional Phrase Attachment Rule 2

Syntactic pattern:        PP, S
Additional restrictions:  none
Syntactic assignment:     PP is attached to S
Semantic action:          Fill the slot (specified by the preposition)
                            of the S with the NP in the PP
Result:                   S
```

Since prepositional phrases must occur after noun phrases which they modify, but can occur either before or after verbs, attachment rules 1 and 2 above are needed.

Often, it is ambiguous syntactically where a prepositional phrase should be attached, and exactly what slot the preposition refers to. For instance, consider the following examples:

English:
John ate the cake with a fork.
John ate the cake with chocolate frosting.

The resolution of ambiguities is handled by the Generalized Syntactic Rule selection process. In these examples, "with" is semantically ambiguous, referring to many possible relations, among which are INSTRUMENT and PART-OF, its two semantic meanings in the above examples. Thus, the definition of "with" contains pointers to these relations, as was discussed in the last chapter. Since "fork" is a type of TOOL, used for eating, the Generalized Syntactic Rule selection process chooses INGEST INSTRUMENT FORK as a desirable connection. The Generalized Syntactic Rule which performs this connection is the Prepositional Phrase Attachment Rule 1, operating on "ate" and "with a fork." Thus, the PP is attached to the verb in the first example. In the second example, on the other hand, semantics would prefer to connect "cake" and "chocolate frosting". Thus, the PP is attached to "cake," and the PART-OF relation is chosen as the meaning of "with."

Similar situations occur in French and Spanish. Here is a French example:

French: Jean mangeait un gateau au chocolat.

English: John was eating a chocolate cake.

French: Jean mangeait un gateau au restaurant.

English: John was eating a cake at the restaurant.

As in the English example, it is ambiguous as to where the prepositional phrase beginning with "au" should be attached. MOPTRANS determines which attachment is appropriate in the same way as the English examples. "Au" is defined to mean PLACE or MADE-OF (among other meanings). In the first example, since "chocolat" is a type of FOOD, as is "gateau" (cake), the semantic rule selection process chooses the attachment (FOOD MADE-OF FOOD) as a possible connection. Then, since Prepositional Phrase Attachment Rule 1 can perform this slot-filling, it is executed on "gateau" and "au chocolat." In the second example, "restaurant" is a BUILDING, which is a LOCATION. Thus, the relation PLACE between INGEST (an ACTION) and LOCATION is preferred by the rule selection process, and Prepositional Phrase Attachment Rule 1 attaches "au restaurant" to "mangeait."

Verbs or nouns can also govern the semantic meaning and attachment of prepositional phrases. This is done by the following rule:

```
Specific PP Meaning Rule:

Syntactic pattern:         NP or S, PP
Additional restrictions:   NP or S expects PREP in PP to refer to a particular
                             slot
Syntactic assignment:      PP is attached to S
Semantic action:           Fill the slot (specified by the NP or S)
                             of the S with the NP in the PP
Result:                    NP or S, NP in PP
```

An example of when this rule is used is with the verb "to search." In English, the OBJECT of the conceptualization built by "search" is found after the preposition "for." This information is encoded in the dictionary definition of "search." Because of this, the Specific PP Meaning Rule is executed whenever a form of "to search" appears in a sentence, followed by the word "for" and a noun phrase which fits the prototype for the OBJECT of the conceptualization built by "search." Similarly, in French the OBJECT of an MTRANS expressed by the verb "penser" (to think) appears after the preposition "a." This information is stored in the dictionary definition of "penser," causing this rule to execute whenever "penser" is followed by "a."

7.2.2.3 Relative Clauses

The relative clause rules used in MOPTRANS are shared among English, Spanish, and French. To illustrate how they work, let us consider the following examples of English relative clauses:

I saw the man who gave the book to Mary.
I saw the book given to the man by Mary.
I saw the book that Mary gave to John.
I saw the man who Mary gave the book to.
I saw the man to whom Mary gave the book.
I saw the man who Mary said gave the book to John.
I saw the man who Mary said John gave the book to.

These examples illustrate the fact that the gap in a relative clause in English can come almost anywhere. In these relative clauses, the gap occurs in the position of the subject of the clause verb ("gave" or "given") in the first and second examples, the direct object of "gave" in the third example, the indirect object of "gave" in the fourth and fifth examples, the subject of the embedded clause verb "gave" in the sixth example, and finally the indirect object of the embedded clause verb in the last example.

For the purposes of explaining MOPTRANS' relative clause rules, we can divide these examples into three separate cases: clauses whose gaps appear as the subject of the clause verb (the first and second examples above), clauses whose gaps appear somewhere after the clause verb (the third, fourth, sixth, and seventh examples above), and clauses in which a preposition appears directly before the relative pronoun (the fifth example above).

Some of the rules which process the first class of relative clauses, those missing a subject, have already been presented. These include the Unmarked Passive Rule, discussed in chapter 6. Here is a list of all of the rules for this class, along with examples of clauses which they are responsible for processing:

```
Unmarked Passive Rule
(for "I saw the book given to the man by Mary.")

Syntactic pattern:         NP, VPP
Additional restrictions:   none
Syntactic assignment:      NP is (syntactic) SUBJECT of VPP, VPP is PASSIVE,
                              VPP is a RELATIVE CLAUSE of NP
Semantic action:           NP is (semantic) OBJECT of S (or another
                              slot, if specified by S)
Result:                    NP, VPP (changed to S)


Participial Phrase Rule
(for "The man giving the book to Mary was seen by me.")

Syntactic pattern:         NP, V
Additional restrictions:   V is present participle
Syntactic assignment:      V is a RELATIVE CLAUSE of NP, NP is SUBJECT
                              of V
Semantic action:           NP is ACTOR of V (or another slot, if specified
                              by V)
Result:                    NP, V (changed to S)
```

Marked Subject-gap Clause Rule
(for "I saw the man who gave the book to Mary.")

Syntactic pattern:	NP, RP (relative pronoun), V
Additional restrictions:	none
Syntactic assignment:	V is a RELATIVE CLAUSE of NP, NP is SUBJECT of V
Semantic action:	NP is ACTOR of V (or another slot, if specified by V)
Result:	NP, V (changed to S)

The second class of relative clauses, those in which the gap appears somewhere after the verb, is somewhat more difficult, due to the fact that the location of the gap must be found. This class is handled by the following rules:

Clause Rule for Gaps After the Verb (CGAV Rule)
(for "I saw the book that Mary gave to John," "I saw the man who Mary gave the book to," "I saw the man who Mary said gave the book to John," and "I saw the man who Mary said John gave the book to.")

Syntactic pattern:	NP, RP (optional), S
Additional restrictions:	none
Syntactic assignment:	S is a RELATIVE CLAUSE of NP
Semantic action:	none
Result:	NP, S (changed to CLAUSE-VERB)

Gap-finding Rule

Syntactic pattern:	NP, CLAUSE-VERB, <anything>
Additional restrictions:	Item following the CLAUSE-VERB must not be a NP, or what follows must not partially match any rules that could lead to the building of an NP
Syntactic assignment:	none
Semantic action:	none
Result:	NP, CLAUSE-VERB, NP, <anything>

Wrong Gap Rule (an error-correction rule)

Syntactic pattern:	NP, CLAUSE-VERB, NP (copy of first NP)
Additional restrictions:	none
Syntactic assignment:	none
Semantic action:	none
Result:	NP, CLAUSE-VERB

These three rules process this class of relative clause in the following way: first, the CGAV Rule marks the clause verb as such. Next, the Gap-finding Rule attempts to find a gap in the clause, by looking for any position after the verb which is not directly followed by a NP. The reason for this restriction is so that the rule does not attempt to fill what it thinks is a gap when in reality an NP which can fill that gap actually exists directly afterward in the text (e.g., in "I saw the man who Mary said John gave the book

to," we know that the gap is not after "said" because it is immediately followed by the NP "John"). When a space is found, the Gap-finding rule places a copy of the NP which dominates the clause in the location which it has found. Then, attachment of the NP is left to the system's regular rules, such as the Object Rule, and the Prepositional Phrase Rule.

To illustrate how these rules work, let us examine in detail the processing of the sentence "I saw the book that Mary gave to John." First, the CGAV Rule marks "gave" as a CLAUSE-VERB, since the S built from "Mary" and "gave" follows the relative pronoun "that." Next, the Gap-finding Rule places a copy of the NP built by "the book" in active memory after "gave," since "to" follows, and "to" could not lead to the building of an NP. At this point, active memory contains the CLAUSE-VERB followed by the NP, "the book." Next, the Object Rule attaches "the book" to "gave," since an S is followed by a NP in active memory, and since "book" is a PHYS-OBJECT, which matches the prototype for the OBJECT of an ATRANS. Thus, the OBJECT gap in the clause is filled.

The Gap-finding Rule does not always place a copy of the NP before the clause in the right place. Because of this, the Wrong Gap Rule is needed to remove the copy of the NP in cases where subsequent processing proves that the gap is further in the sentence. The Wrong Gap Rule is used to process the sentence "I saw the man who Mary said John gave the book to." In this sentence, the Gap-finding Rule incorrectly identifies what it thinks is a gap several times before the correct gap location is found. First, it identifies a possible location for a gap after "John," since this is after the clause verb "said," and a NP does not follow. Thus, it inserts a copy of the NP built by "the man" in active memory after "John," leaving active memory with the CLAUSE-VERB, followed by two NP's. After the Gap-finding Rule finishes, no rules match this pattern, and thus the Wrong Gap Rule is executed, removing the just-inserted copy of "the man" from active memory. Next, the Gap-finding Rule tries again, this time identifying a potential gap after "book." Again, a copy of "the man" is inserted, but no rules can match the resulting pattern in active memory. Thus, the Wrong Gap Rule removes the copy from active memory. Finally, the Gap-finding Rule correctly identifies the gap after the word "to," and inserts a copy of "the man" in this location. This leads to the execution of the Prepositional Phrase rule, attaching "the man" as the object of the preposition "to," which in turn leads to the execution of the Indirect Object rule, which assigns "the man" as the RECIPIENT of the ATRANS of the book by Mary.

The third class of relative clauses, in which a preposition appears before the relative pronoun, is processed by the following rules:

```
Preposition-marked Relative Clause Rule (PRC Rule)

Syntactic pattern:          NP, PREP, RP, S
Additional restrictions:    none
Syntactic assignment:       NP is the PREP-OBJECT of PREP,
                            PREP is the RC-PP of NP (PREP is changed to a PP),
                            S is a REL-CLAUSE of NP
Semantic action:            NP will be filler of the slot specified by the
                              PREP in some conceptualization
Result:                     NP, S (changed to a PP-CLAUSE-VERB)
```

PP Gap-finding Rule

Syntactic pattern:	NP, V
Additional restrictions:	NP has a RC-PP
Syntactic assignment:	none
Semantic action:	none
Result:	NP, V, PP (the RC-PP of NP)

Wrong PP Gap Rule (an error-correction rule)

Syntactic pattern:	NP, V, PP
Additional restrictions:	PP is the RC-PP of NP
Syntactic assignment:	none
Semantic action:	none
Result:	NP, V

These rules work in much the same way as the rules for the second class of relative clauses. The PP Gap-finding Rule finds a verb which the PP could possibly attach to, and inserts the PP in active memory. Then, the normal PP attachment rules attach the PP, if it makes sense semantically. Otherwise, the Wrong PP Gap Rule removes the PP, and the PP Gap-finding Rule finds a V later in the clause.

To illustrate how this works, consider the following two examples:

I saw the man to whom Mary gave the book.
I saw the man about whom Mary wanted to know more.

In the first example, the PP Gap-finding Rule places the PP "to the man," which it built during execution of the PRC Rule, after "gave." Then, Prepositional Phrase Attachment Rule 1 attaches this PP to "gave," assigning "the man" as the RECIPIENT of the ATRANS. In the second example, the PP Gap-finding Rule first places the PP "about the man" after "wanted." However, this time, no PP attachment rules succeed in attaching this PP to "wanted," since semantically this attachment does not make sense. Thus, the Wrong PP Gap Rule removes the PP. Then, the PP Gap-finding Rule tries again, placing the PP after "know." This time, Prepositional Phrase Attachment Rule 1 succeeds in attaching "about the man" to "know," thus assigning "the man" as the OBJECT of the MTRANS built by "know."

These relative clause rules enable the MOPTRANS parser to parse a wide variety of relative clauses. The nesting of the gap in the clause can be arbitrarily deep, as is illustrated by the example "I saw the man who Mary said John gave the book to." However, there are some examples for which this set of rules is not sufficient, as the following examples illustrate:

The secretary who the boss wanted to type a letter was on her lunch break.

The old woman who the boy scout wanted to help across the street hit him with her purse.

In the first sentence, the gap in the clause occurs after "wanted," whic is the first place that the Gap-finding Rule tries. Thus, MOPTRANS can parse this sentence correctly with the rules described above. However, in the second sentence, the gap in the clause is after "help," not after "wanted." The Gap-finding Rule first attempts to identify the gap as after "wanted," and places a copy of the representation of "the old woman"

after "wanted." When "to help" is read, the rule which attaches the subject of an infinitive phrase to the infinitive is executed, since "the old woman" fits the prototype for the ACTOR slot of "help." Thus, with the rules I have presented, the location of the gap would be incorrectly identified, and the sentence would be misparsed.

To handle examples such as this, we can add backup rules telling MOPTRANS when it has incorrectly identified a gap. The backup rule for this example would be the following:

```
Gap-filling Backup Rule
(for "The old woman who the boy scout wanted to help across the street hit
      him with her purse.")

Syntactic pattern:         V, <anything>
Additional restrictions:   V is transitive, what follows the verb
                             cannot partially match any rules which will
                             build an NP
Action:                    Back up to the execution of the Gap-finding Rule
```

To process the above example, the Gap-filling Backup Rule would be executed after MOPTRANS encountered "across," which cannot match any rules to build an NP after "help," which is a transitive verb. Then, MOPTRANS would back up to the execution of the Gap-finding Rule which placed a copy of "the old woman" after "wanted." The Gap-finding Rule would be prohibited from re-executing immediately, and instead the next word would be processed. Then, "the old woman" would be placed after "help" by the Gap-finding Rule, and "the old woman" would be attached to "help" as its OBJECT by the following rule:

```
"Help" rule

Syntactic pattern:         a form of "help", V
Additional restrictions:   V must be tenseless
Syntactic assignment:      V is the INF-CLAUSE of "help"
Semantic action:           V is the OBJECT of "help"
Result:                    NP, S (changed to a PP-CLAUSE-VERB)
```

After the execution of this rule, the parse would continue correctly.

The above relative clause rules also enable MOPTRANS to parse a wide variety of clause constructions in Spanish and French. Here are some examples of Spanish and French clause constructions, as well as the rules which would process them:

Spanish: Yo vi al hombre que dio el libro a Maria.

French: J'ai vu l'homme qui a donne le livre a Marie.

(I saw the man who gave the book to Mary.)

Rule: Marked Subject-gap Clause Rule

Spanish: Yo vi el libro que Maria le dio a Juan.

French: J'ai vu le livre que Marie a donne a Jean.

(I saw the book that Mary gave to John.)

Rules: CGAV Rule, Object Rule

Spanish: Yo vi el hombre a quien Maria dio el libro.

French: J'ai vu la personne a qui Marie a donne le livre.

(I saw the man who Mary gave the book to.)

Rules: PRC Rule, Prepositional Phrase Attachment Rule 1

Spanish: Yo vi a la persona que Maria me dijo le dio el libro a Juan[13].

French: J'ai vu la personne qui Marie a-t-elle dit a donne le livre a Jean.

(I saw the person who Mary said gave the book to John.)

Rules: CGAV Rule, Subject Rule

7.3 Language-Specific Rules

7.3.1 English Noun Groups

In English, the lack of morphological markings on the words makes the parsi noun groups more difficult than in MOPTRANS' other languages. Many verbs ca function as nouns or adjectives with no spelling change. Past active verbs oft function as past participles or adjectives. Nouns can also be used as adjectives. one of the problems in English parsing is identifying the part of speech of the wo the sentence.

Because of this, it is important in English to identify noun group boundaries processing the words within a noun group, since the location of syntactic boundari often play a key role in determining the syntactic role, and even the meaning, of within a noun group. To illustrate this, consider the following noun group:

The red car seat

"Car" functions as an adjective in this example, since the end of the noun grou not come right after "car" but after "seat." Because "car" functions as an adjecti adjective "red" does not apply to it, but rather to "seat," the head noun of the group.

To identify the head noun of the noun group, the MOPTRANS parser us following *Generalized Syntactic Rule:*

```
Head Noun Rule

Syntactic pattern:          N, <any word>
Additional restrictions:    The word following the N is not a N.
Syntactic assignment:       none
Semantic action:            none
Result:                     N (changed to HN), <any word>
```

[13]This sentence sounds very awkward or even ungrammatical to some native Spanish, due to th of the gap *inside of the second clause.*

Things are not always this easy, however, since it is not always clear as to whether or not the next word is a noun. Consider this example, discussed in (Gershman, 1979):

The U.S. forces fight in Vietnam is hopeless.

Since either "forces" or "fight" could be verbs, the decision as to which word is the head noun is not straightforward. The approach to this problem that Gershman's parser took was to collect as many words as possible in a noun group. Thus, in this example, "forces" and "fight" would be assumed to be nouns, and collected in the noun group. This approach does not work for the following example, however:

The U.S. forces fight in Vietnam.

Here, since "fight" is a verb, backup would be required. Fortunately, in this example, the meaning of the sentence does not depend on the syntactic class of the word "fight," since the noun "fight" and the verb "fight" mean the same thing. However, there are examples for which this is not true:

Mickey Mouse watches people at Disneyland buy are expensive.
Mickey Mouse watches people at Disneyland to see if they behave.

To handle examples like these, it seems clear that the parser must back up in one of the two cases, since the point at which it is certain whether "watches" is a noun or a verb is very late in the sentence (at "buy" in the first example, and "to see" in the second example). Thus, the MOPTRANS parser sometimes uses a backup rule for noun groups:

```
Noun Group Backup Rule

Syntactic pattern:          HN
Additional restrictions:    HN could have been a V, no V appears
                               later in the sentence
Action:                     Back up to the execution of the Head Noun Rule.
```

At the point at which the infinitive phrase is encountered in the second example, the backup rule knows that no main verb will be encountered. Thus, backup occurs, and "watches" is assumed to be a verb in the second example.

7.3.3 German Parsing Rules

The German language allows freer word order than do English, Spanish, or French. Since noun phrases carry case markings, this information can provide clues as to what is the subject or direct object of a verb which must be provided by word order in the other languages. Thus, it is grammatical in German to order a sentence SOV instead of SVO. At times, such as in relative clauses, the SOV ordering is required. At other times, verbs are separated, so that a part of the verb or an auxiliary verb comes between the subject and direct object, and the remainder of the verb appears at the end of the clause.

Because of the freer word order of German, and the case markings provided in the languages, the MOPTRANS parser's rules for parsing German are not as similar to English, French, and Spanish as the rules for these languages are to each other. Parsing German relative clauses, sentences with auxiliary verbs, and other constructions requires a different set of rules.

Noun Group Attachment

Since case markings are more important and word order is less important in German, the attachment of subject and direct object to the verb is more like prepositional phrase attachment in English than it is to the English subject and direct object rules. The case provides some information as to what relation exists between the verb and the noun group, just as a preposition provides information as to what semantic relation exists between its object and the constituent which it attaches to.

Because of this, the German prepositional phrase attachment rules and subject and object attachment rules are all the same. To completely unify their attachment process, prepositions are treated as case assignments by the prepositional phrase rule:

```
German PP Rule

Syntactic pattern:         PREP, NP
Additional restrictions:   none
Syntactic assignment:      Assign the PREP as the CASE of the NP
Semantic action:           none
Result:                    NP
```

Thus, a prepositional phrase is treated as a noun group, with a different case marking. A regular noun group has a case like NOMINATIVE, ACCUSATIVE, DATIVE, or GENITIVE. However, the preposition is assigned to be the case of a noun group which came from a prepositional phrase. Thus, these noun groups would be marked by cases such as "von" (by), "nach" (to), etc.

Attachment of noun groups to verbs is handled in all cases by the following rules:

```
German NP Before V Rule

Syntactic pattern:         NP, V
Additional restrictions:   none
Syntactic assignment:      Assign the NP as the case-filler of the V,
                              according to the case of the NP
Semantic action:           If the V specifies that the case of the NP
                              should fill a particular slot, then fill
                              that slot with the NP. Otherwise, perform
                              the default slot-filling associated with the
                              NP's case.
Result:                    V
```

German V Before NP Rule

Syntactic pattern:	V, NP
Additional restrictions:	none
Syntactic assignment:	Assign the NP as the case-filler of the V, according to the case of the NP
Semantic action:	If the V specifies that the case of the NP should fill a particular slot, then fill that slot with the NP. Otherwise, perform the default slot-filling associated with the NP's case.
Result:	V, NP

Verbs can specify what particular slot should be filled by a noun group with a particular case. This is true whether the case is morphologically marked, such as NOMINATIVE or DATIVE, or marked by a preposition. Thus, the verb "empfangen" (received) provides the information that the NOMINATIVE case fills the RECIPIENT slot of the ATRANS built by the verb, rather than the ACTOR slot, which is the default slot-filling for the NOMINATIVE case. Similarly, since the verb "warten" (wait) expects to find the preposition "auf" (for) following it, followed by its semantic OBJECT, "warten" provides the information that a noun group marked with the "auf" case fills the OBJECT slot.

Case Markings

German cases are sometimes ambiguous; that is, one often cannot tell whether a noun is marked as one case or another. Often semantics can resolve case ambiguities; in fact, sometimes semantic considerations are essential to the determination of case. Consider the following example, which was encountered by MOPTRANS:

German: Iran sagte heute dass irakische Agenten waehrend eines Angriffes in der Naehe von der irakischen Grenze 2 Maenner toeteten und mehrere Geisel nahmen.

English: Iran today said iraqi agents killed two men and seized a number of hostages in a raid near the border with iraq.

After the word "dass," the German equivalent of "that," the verb must come at the end of the clause that follows. In this case, the clause is a conjunctive clause. The verb in the second portion of the conjunction is "nahmen" (took or seized). Since the conjunction could either conjoin two verb phrases with the same subject, or two sentences, "Geisel" (hostage) syntactically could either be nominative or accusative, functioning as the subject or the direct object of "nahmen." Semantics resolves this ambiguity, because the concept HOSTAGE, which represents the noun "Geisel," prefers to fit into the OBJECT slot of GET-CONTROL, the representation of "nahmen." Thus, the semantic preference results in the choice of the ACCUSATIVE case for "Geisel," and "Geisel" is assigned as the syntactic direct object of "nahmen."

There are cases where semantics does not provide enough information to disambiguate the case of a noun group. Because of this, the MOPTRANS parser also relies on word order information to choose the case of a noun group. By default, if

semantic considerations do not suggest otherwise, the first noun group in a sentence is considered to be nominative if the case marking of that noun group is ambiguous (which it usually is). For example, consider the following sentence:

German: Das Buch das John Mary gab war interessant.

English: The book that John gave Mary was interesting.

In this sentence, it is not distinguishable from morphological information whether "John" and "Mary" are nominative or accusative. There are also no semantic preferences as to whether "John" or "Mary" should be the ACTOR or the RECIPIENT of the ATRANS. To handle situations like this, the MOPTRANS parser assumes by default that the first noun group in a list of noun groups is nominative. Thus, MOPTRANS parses the above sentence as (ATRANS ACTOR JOHN RECIPIENT MARY OBJECT BOOK).

Subclauses

In German, the verb in a subclause must come at the end of the clause. As I discussed earlier, often it is the case that before the verb of a subclause is reached, the action to which the verb refers can be inferred. This is because the slot-fillers specified by the noun groups and the semantic roles which they play as specified by their case markings sometimes provide enough information to infer the action which they are playing a role in.

In order to anticipate the verb of a clause, the MOPTRANS parser builds a "dummy" action whenever it encounters the beginning of a clause. Then, the same noun group attachment rules as used above can be used to attach noun groups or prepositional phrases to the dummy action. This allows the refinement process to infer the action before encountering the verb, if the noun groups or prepositional phrases provide specific enough role-filling information to allow this inference to occur. The rules which allow this to happen are the following:

```
German Relative Pronoun Rule

Syntactic pattern:          RP
Additional restrictions:    none
Syntactic assignment:       RP is a RELATIVE CLAUSE
Semantic action:            Build the concept ACTION
Result:                     RP (changed to V)

German Clause Verb Rule

Syntactic pattern:          V, V
Additional restrictions:    V1 is a RELATIVE CLAUSE
Syntactic assignment:       none
Semantic action:            Merge the representation for V2 with V1
Result:                     V1
```

Thus, once a relative pronoun has been found, the MOPTRANS parser treats it as though it were a verb, so that NP's and prepositional phrases can be attached to it, and the concept refinement demons can infer the particular action that the relative clause

must refer to, if the NP's and PP's provide enough information.

Auxiliary Verbs

A similar phenomenon in German is the separation of the verb into two non-adjacent words. This occurs when auxiliary verbs are used. The auxiliary verb is placed where the verb normally appears, between the subject and direct object, and the participial form of the verb appears at the end of the clause. For instance, here is a passive sentence encountered by MOPTRANS:

> German: Schwarzzivilrechtsfuehrer Vernon Jordan wurde am Donnerstag in einem Motelparkiergrund von einem unidentifizierten Schuetzen in dem Ruecken geschossen.
>
> English: Black civil rights leader Vernon Jordan was ambushed and shot in the back by an unidentified sniper in a motel parking lot Thursday.

"Wurde," the German passive auxiliary, appears after the subject, "Scharzzivilrechtsfuehrer" (Black civil rights leader), where the verb normally appears in the sentence. The past participle, "geschossen" (shot), is placed at the end of the sentence.

Again, it is sometimes possible to infer the action in the sentence before encountering the past participle. Thus, the MOPTRANS parser builds a "dummy" action for auxiliary verbs, similar to the way in which clauses are handled.

German auxiliaries are handled in MOPTRANS by the following rules:

```
German Auxiliary Verb Rule

Syntactic pattern:         AUX
Additional restrictions:   none
Syntactic assignment:      none
Semantic action:           Build the concept ACTION
Result:                    RP (changed to V)

German Auxiliary Attachment Rule

Syntactic pattern:         V, V
Additional restrictions:   V1 is an AUX, V2 is a (past or present) participial
Syntactic assignment:      none
Semantic action:           Merge the representation for V2 with V1
Result:                    V1
```

In passive constructions, the noun group which would normally be nominative in an active sentence is marked by by the preposition "von," as in the example above ("von einem unidentifizierten Schuetzen" is equivalent to "by an unidentified sniper"). However, the preposition "von" can also mean "from." Thus, the appearance of "von" after the passive auxiliary is potentially ambiguous, as the following example illustrates:

> German: Der Mann wurde in einem Wagen von Hotel gefuehrt.
>
> English: The man was driven from the hotel in a car.

Here, the appearance of "von" after "wurde" does not indicate the ACTOR, as it does when it is equivalent to the English passive "by." Instead, it refers to the FROM slot of the PTRANS.

To handle this ambiguity, the MOPTRANS parser uses the following rule:

```
German Passive Subject Rule

Syntactic pattern:          V, NP
Additional restrictions:    V is passive auxiliary (a form of "wurden"),
                               the case of NP is "von"
Syntactic assignment:       NP fills the "von" case of V
Semantic action:            If NP is a LOCATION, then NP is the LOCATION of V
Result:                     V, NP
```

If "von" does mark the subject of the passive, then the noun group is attached syntactically, under the "von" case of the V. It is not attached semantically in this situation, because the V could specify that its subject should fill a slot other than ACTOR. Thus, the slot in which the subject should be placed cannot be definitely known until the past participle is found. As a result, one of the actions which is performed in the Auxiliary Attachment Rule is the assignment of the "von" case of the auxiliary as the SUBJECT of the participial, if the auxiliary is passive, and the "von" case NP was not a LOCATION.

Separable Prefixes

Some German verbs are separated, similarly to the way in which auxiliaries are separated from their participles, whenever they appear in conjugated form. These verbs are called *separable prefix verbs*. These verbs are made up of a stem, which is itself an infinitive; and a prefix, which changes the meaning of the stem when it is placed onto the beginning of it.

For example, the German infinitive "ueberfallen," which means "to assault," is made up of a stem "fallen," the German infinitive meaning to fall or drop, and the prefix "ueber." When infinitives like this are used in a conjugated form, the stem and the prefix are split, and appear in different parts of the sentence. The stem appears in the same position in the sentence where the verb normally appears, and the prefix appears at the end of the clause. Thus, the prefix of "ueberfallen" is separated in the following example:

German: Vermutete baskische Guerrillen fiel 2 Polizeiwagen am Donnerstag Nacht mit Sprengstoff ueber und verwundeten 6 Polizisten.

English: Presumed Basque separatist guerrillas ambushed two national police cars with explosives thursday night, wounding six policemen.

Here, "fiel" is the third person past tense of "fallen," the stem of "ueberfallen," and "ueber" appears at the end of the clause in which "ueberfallen" is the main verb.

To handle separable prefixes, the MOPTRANS parser has the following Generalized Syntactic Rule:

German Separable Prefix Rule

```
Syntactic pattern:          V, PREFIX
Additional restrictions:    PREFIX is a separable prefix with stem V
Syntactic assignment:       none
Semantic action:            Change the semantic representation of V to the
                               meaning of the separable infinitive
Result:                     V
```

As is the case when verbs appear at the end of a clause, it is sometimes possible with separable prefixes to infer the separated infinitive before the prefix is encountered in the sentence. This is the case in the sentence above, since the subject of "fiel" is "Guerrillen" (guerrillas), and the prepositional phrase "mit Sprengstoff" (with explosives) appears after the verb stem. These cues give a native German speaker enough information to infer that the infinitive is "ueberfallen."

The way in which the MOPTRANS parser parses separable prefixes also allows it to infer the infinitive at times before the prefix is encountered. To do this, the definition of the stem of a separable prefix points to its default meaning, along with all the meanings of separable infinitives containing that stem. Thus, when a stem is first parsed, a dummy structure is built as its representation. During the parse, the concept refinement demons can choose the stem's default meaning, or one of the meanings of the separable infinitives, if any slot-fillings provide enough information for the demons to do so.

In the example above, "fallen" is defined with pointers to its default definition, fall (PTRANS DIRECTION DOWN), as well as ASSAULT, the definition of "ueberfallen," and other separable prefix infinitives in the system which have "fallen" as their stem. Since the context of the example above causes GUERRILLA to be filled in as the ACTOR of "fallen" and BOMB as the INSTRUMENT of "fallen," this slot-filling information is enough to cause the semantic refinement process to refine "fallen" to the more specific concept ASSAULT. Thus, when the parser reaches the prefix "ueber" and executes the separable prefix rule, the representation is not changed at all, since ASSAULT has already been inferred.

7.3.4 Chinese

Chinese is, for the most part, an SVO language. For simple sentences, the subject precedes the verb, which is followed by the direct object:

Chinese: Wo mae le i ben shu.

Literal English: I buy (past marker) one (classifier) book.

English: I bought a book.

"Le" in Chinese is more or less a past tense marker. "Ben," and other classifiers, are used after numbers, to mark them as adjectives. The particular classifier used depends on the semantics of the object being counted (in this case, "book".)

Although the SVO word ordering of Chinese is similar to Indo-European languages, verb phrases can also function similarly to prepositional phrases. For instance, consider the following sentence:

Chinese: Wo tzuo tzai juotz shang.

Literal English: I sit am-at table on-top-of.

English: I sit on the table.

The word "tzai" is a verb, meaning "to be at." Here, it corresponds to the English preposition "on" or "at," because it follows another verb. The post-position, "shang," meaning "on top of," further specifies the meaning of the prepositional phrase, to mean "at the top of," or "on."

Verb phrases functioning as prepositional phrases can also come before the main verb of the sentence in Chinese, as in the following:

Chinese: Wo gei JangSan mae le i ben shu.

Literal English: I gave John bought (past marker) one (classifier) book.

English: I bought John a book.

In this example, "gei" (to give) functions as a preposition, since the verb "mae" (bought) is marked with the word "le" as being the main verb of the sentence. When "gei" is used as a preposition in Chinese, it usually marks the recipient of the main action. Thus, in this sentence, it is equivalent to the English "to," the indirect object marker.

Thus, a verb phrase in Chinese can play two roles, and these roles determine how it can be attached to other parts of the sentence. A verb phrase can always be attached to a noun phrase, whether it is the main verb of the sentence, or functioning as the equivalent of a prepositional phrase. However, when it is the main verb, it cannot attach to another verb, while attachment to another verb is legal if it is functioning as a prepositional phrase.

To handle verb phrases in Chinese, the following rules are used in the MOPTRANS parser:

```
Chinese Verb Phrase Rule

Syntactic pattern:         V, NP
Additional restrictions:   none
Syntactic assignment:      NP is the (syntactic) OBJECT of the V
Semantic action:           NP is the (semantic) OBJECT of the V (or some
                              other slot, if specified by V)
Result:                    V (changed to a VP)

Chinese VP Attachment Rule 1

Syntactic pattern:         NP, MVP
Additional restrictions:   none
Syntactic assignment:      NP is the (syntactic) SUBJECT of the MVP
Semantic action:           NP is the (semantic) ACTOR of the MVP (or some
                              other slot, if specified by V)
Result:                    MVP
```

```
Chinese VP Attachment Rule 2

Syntactic pattern:          VP, MVP
Additional restrictions:    none
Syntactic assignment:       VP is attached to MVP
Semantic action:            VP fills the slot of MVP as specified by V
                               in VP
Result:                     MVP

Chinese Main VP Rule

Syntactic pattern:          VP, "le" (or some other particle)
Additional restrictions:    none
Syntactic assignment:       none
Semantic action:            none
Result:                     VP (changed to MVP)
```

More than one main VP may be found in a sentence, as was illustrated in an example discussed earlier:

> Chinese: yilang jintian shuo , yilake tewu xiji yilake bianjing, dasi le er ren , zhuazou le xuduo renzhi.
>
> Literal English: Iran today say, iraqi agents attack iraqi border, kill (past marker) 2 men, seize (past marker) a number of hostages.
>
> Good English: Iran today said iraqi agents killed two men and seized a number of hostages in an attack on the iraqi border.

This construction functions as does conjunction in English between verb phrases. To handle the stringing together of verb phrases, MOPTRANS uses the following rule:

```
Chinese VP Attachment Rule 3

Syntactic pattern:          MVP, MVP
Additional restrictions:    none
Syntactic assignment:       MVP2 is attached to MVP1,
                            NP of MVP1 is the (syntactic) SUBJECT of the MVP2
Semantic action:            NP of MVP1 is the (semantic) ACTOR of MVP2 (or some
                               other slot, if specified by the V)
Result:                     MVP2
```

7.3.5 Relative Clauses

In the case of the sentence "Wo gei JangSan mae le i ben shu" (I bought John a book), these rules work as follows: first, the Chinese VP Rule forms a VP from "gei JangSan" (gave John). Then, since "mae" (bought) is marked with "le," both the VP Rule and the Main VP Rule apply to "mae le i ben shu" (bought a book). Finally, VP Attachment Rule 2 attaches "gei JangSan" to "mae le i ben shu," filling the RECIPIENT slot of "mae" (BUY), due to information stored in the dictionary definition of "mae" which says that it can refer to the RECIPIENT slot.

Dependent clauses come before the noun phrase that they modify in Chinese. Here is an example:

Chinese: Wo mae de shu ...

Literal English: I bought (particle) book ...

English: The book I bought ...

The particle "de" marks the verb phrase as being a relative clause, rather than the main clause of the sentence. However, even though the verb phrase is not the main clause of the sentence, it is the main verb phrase of the relative clause, and therefore other verb phrases can attach to it as prepositional phrases.

MOPTRANS' rules for relative clauses in Chinese are the following:

```
Chinese Relative Clause Marker Rule

Syntactic pattern:        VP, "de"
Additional restrictions:  none
Syntactic assignment:     none
Semantic action:          none
Result:                   VP (changed to CL)

Chinese Relative Clause Rule

Syntactic pattern:        CL, NP
Additional restrictions:  none
Syntactic assignment:     CL is attached to NP
Semantic action:          NP fills a slot of CL (choose a slot which
                             semantically qualifies)
Result:                   NP
```

7.4 Conclusion

We have seen that a great deal of the parsing knowledge in MOPTRANS is applicable to more than one language. MOPTRANS' conceptual knowledge base, which makes up a large percentage of the system's total knowledge, is used unchanged to parse all of the system's languages. Likewise, many of the Prototype Failure Rules are used unchanged for all five languages. Finally, there are even some syntactic rules which are used for all five languages, including the Conjunction Rule and pronoun resolution strategies.

The ability to share knowledge in MOPTRANS is important for three reasons. First, it reflects the large amount of non-linguistic, conceptual knowledge that must be used in parsing. This knowledge, since it is conceptual, is language-independent. Thus, the same knowledge is applicable to the processing of any language. This is reflected in MOPTRANS by the fact that the same conceptual knowledge base is used to parse all five of the system's input languages.

Second, the shared syntactic knowledge in MOPTRANS reflects the similarities in languages of the same families. Intuitively, English, Spanish, and French seem similar to

each other, and MOPTRANS' body of syntactic rules provides evidence to confirm this intuition. On the other hand, Chinese, a non-Indo-European language, shares vastly fewer rules with other languages in the system, reflecting its more distant relationship to the other languages.

Finally, the sharing of knowledge in MOPTRANS allows for an efficiency of representation that would not be possible in a request-based parser. Since requests contain conceptual knowledge to guide the search for slot-fillers, in addition to syntactic knowledge about where to look for these fillers, conceptual knowledge must be duplicated in requests for every language in the system. For example, consider the following sentences:

Iran seized control of the U.S. embassy.
A gunman seized control of a Boeing 727 and diverted it to Cuba.
The embassy seized by Iranian students was American.
The plane seized by a gunman was diverted to Cuba.

In these examples, "seized" is ambiguous, both syntactically and semantically. It can be a past active or past participle, and it can refer to the structures TAKE-OVER-BUILDING or HIJACK. To disambiguate "seized," the following requests would be needed:

Past active "seized" meaning TAKE-OVER-BUILDING: If a BUILDING appears to the right of "seized," it means TAKE-OVER-BUILDING, and the noun group to the left of the verb is the ACTOR.

Past active "seized" meaning HIJACK: If a VEHICLE appears to the right of "seized," it means HIJACK, and the noun group to the left of the verb is the ACTOR.

Unmarked passive "seized" meaning TAKE-OVER-BUILDING: If a BUILDING appears to the left of "seized," it means TAKE-OVER-BUILDING, and the noun group after "by" is the ACTOR.

Unmarked passive "seized" meaning HIJACK: If a VEHICLE appears to the left of "seized," it means HIJACK, and the noun group after "by" is the ACTOR.

This requests contain semantic knowledge, about what should be the OBJECT of a HIJACK and a TAKE-OVER-BUILDING. Because this knowledge is conceptual, we would like it to be usable in the parsing of other languages, also. However, these rules cannot apply to other languages, because of the syntactic information in them. Thus, to disambiguate "uebernahmen" (to seize or overtake) in German, for example, an entirely different set of requests would be needed, even though the same semantic information is relevant to the disambiguation. Here are the equivalent German sentences:

English: Iran seized control of the U.S. embassy.

German: Iran uebernahm die amerikanische Botschaft.

English: A gunman seized control of a Boeing 727 and diverted it to Cuba.

German: Ein Bewaffneter uebernahm einer Boeing 727 und lenkte sie nach Kuba ab.

English: The embassy seized by Iranian students was American.

German: Die von iranischen Studenten uebernommenen Botschaft war amerikanisch.

Literal English: The by Iranian students seized embassy was American.

English: The plane seized by a gunman was diverted to Cuba.

German: Das von einem Bewaffneten uebernommenen Flugzeug wurde nach Kuba abgelenkt.

Literal English: The by a gunman seized plane was to Cuba diverted.

The semantic information contained in the English requests would have to be duplicated in the German set of requests, combined this time with information about German syntax:

"Uebernahm" meaning TAKE-OVER-BUILDING: If a BUILDING appears to the right of "uebernahm," it means TAKE-OVER-BUILDING, and the noun group to the left of the verb is the ACTOR.

"Uebernahm" meaning HIJACK: If a VEHICLE appears to the right of "uebernahm," it means HIJACK, and the noun group to the left of the verb is the ACTOR.

"Uebernommenen" meaning TAKE-OVER-BUILDING: If a BUILDING appears to the right of "uebernommenen," it means TAKE-OVER-BUILDING, and the noun group to the left of the verb, appearing after "von," is the ACTOR.

"Uebernommenen" meaning HIJACK: If a VEHICLE appears to the right of "uebernommenen," it means HIJACK, and the noun group to the left of the verb, appearing after "von," is the ACTOR.

With separate conceptual and syntactic knowledge in the MOPTRANS parser, however, the same semantic knowledge is used for English and German. The concept refinement rules perform the semantic disambiguation of "seize" and "uebernahmen" in the same way, relying on the following hierarchy:

```
          GET-CONTROL
         /           \
    HIJACK         TAKE-OVER-BUILDING
```

"Seized" and "uebernahmen" are both defined as a GET-CONTROL. Since the OBJECT slot of HIJACK should be filled by a VEHICLE, but the OBJECT slot of TAKE-OVER-BUILDING should be filled by a BUILDING, the Slot-filler Specialization demon chooses one of the two structures when the OBJECT of GET-CONTROL is filled in. The different syntactic rules of English and German cause the parser to fill the OBJECT slot, depending on the constructions of the sentence.

8. Conclusion

8.1 A Short Review

I have presented an approach to integrated parsing in this thesis which differs from previous integrated parsers in some ways, but preserves the following characteristics of integrated parsing:

- Syntactic and semantic processing of a text take place at the same time.
- Syntactic decisions are made with full access to semantic processing; that is, communication between syntax and semantics is high.

As with previous integrated parsers, the motivation behind an integrated approach to natural language processing is to avoid the difficulties in resolving ambiguities in syntax-first parsers. In the context of machine translation, we saw examples of syntax-based transfer rules which attempted to perform word sense disambiguation. These rules were inadequate, because they were rules about a particular lexical item and a particular syntactic construction using that lexical item. Thus, the number of rules needed to handle a large number of cases would be very high. For example, the transfer rule for "realizar diligencias" meaning POLICE-INVESTIGATION was the following:

> If "realizar diligencias" appears in a sentence, its subject has the semantic feature +authority, it is followed by a prepositional phrase consisting of "para" followed by an infinitive with the semantic feature +capture, and the direct object of this infinitive has the semantic feature +criminal, then translate "realizar diligencias" as "to investigate."

This rule was not applicable to similar contexts using the phrase "realizar diligencias," because it relied on the appearance of so many items in the surrounding context in just the right syntactic location. Any change in the syntactic construction of any of these items would require another transfer rule.

Although the goal of integrated processing is the same in MOPTRANS as in previous integrated parsers, the integration of MOPTRANS differs from these previous parsers in the following ways:

- A limited amount of syntactic representation is built during text understanding.
- Knowledge about syntax and semantics is largely separate. Syntactic knowledge is expressed in the parser's knowledge base as a largely separate body of knowledge, but this knowledge has references to semantics, telling the system how semantic representations are be built from these syntactic rules.
- Semantics guides the parsing process, but relies on syntactic rules to make sure that it does not make mistakes.

I have shown that the way in which the MOPTRANS parser is integrated has several advantages over past integrated parsing approaches. They are the following:

Frame selection, or word sense disambiguation

In previous integrated parsers, conceptual and syntactic knowledge were mi together in the same parsing rules. As a result, frame selection knowledge could not expressed without mixing in syntactic knowledge.

Thus, general frame selection rules, such as those used in MOPTRANS, could not encoded in request-based parsers. For instance, in MOPTRANS, the Slot-fi Specialization Rule told the parser when the filling of a particular slot meant that frame whose slot was filled could be refined. In request-based parsers, though, the ana of the Slot-filler Specialization Rule was exemplified by the following request, wł disambiguated the word "seized" to mean TAKE-OVER-BUILDING:

> Look to the right of "seized" for a word which means BUILDING. If such a word is found, build the concept TAKE-OVER-BUILDING, and fill the OBJECT slot of TAKE-OVER-BUILDING with the BUILDING.

This sort of request is far less general than the Slot-filler Specialization Rule, beca it is about a particular *lexical item* ("seized") and a particular syntactic construct (active use of "seized").

Parsing Complex Syntactic Constructions

Previous integrated parsers have attempted to rely on "local" syntactic cues determine the correct syntactic interpretation of texts. This is not adequate for par complex syntactic constructions. In chapter 4, I showed an example of a class of wo verbs which could either be past active or past participle, for which the request-ba approach is not adequate. A large number of complex requests was needed to hai examples of sentences which contained two such verbs, such as the following sentence:

> The soldier called to the sergeant shot in the arm.

Requests for disambiguating "called" also had to disambiguate "shot" before t could determine whether "soldier" was the RECIPIENT or the ACTOR of the MTR. built by "called." Thus, "shot" required a set of special requests just for the situatio which another syntactically ambiguous verb appeared in the sentence before it. Even complex set of requests did not work for the following sentence:

> The soldier called to the sergeant shot in the arm was reprimanded.

Using MOPTRANS' Generalized Syntactic Rules, however, a *simple set of rules* capable of processing any sentence of this type, containing more than one verb w could be past active or past participle. The rules consisted of the Subject Rule, w assigned the noun group before the verb to be its syntactic subject and semantic ACT (or some other slot, if the verb specified); the Unmarked Passive Rule, which assigned noun group before the verb to be its semantic OBJECT; and two backup rules, w were executed if the parser later discovered that an incorrect syntactic assignment been made. These rules were even capable of handling the last example above, wit additional rules needed.

Multi-lingual Parsing

In chapter 7, I discussed the parsing rules which the MOPTRANS parser was able to use for more than one language. This included a large number of different types of rules. All of MOPTRANS' conceptual knowledge was applicable to all of the system's languages.

Even some syntactic rules, such as the Conjunction Rule, and pronominal reference strategies, applied to all of the system's languages. Other syntactic knowledge, such as subject- and object-attachment rules and prepositional phrase rules, applied without changes to English, French, and Spanish, due to the similarities in the syntax of these languages. None of this sharing of knowledge would be possible in previous integrated parsers, due to the intermixing of conceptual and syntactic knowledge in these parsers' rules.

Learning

Although this thesis did not present a theory of language learning, the organization of MOPTRANS' parsing knowledge is more amenable to learning than previous integrated parsers. In a learning system, it is important to store knowledge at as general a level as possible. If this is done, then knowledge learned in one situation can be applied to other, similar situations in which this knowledge is relevant. If this is not done, then the same knowledge must be re-learned many times, because the system cannot determine the range of contexts in which a piece of knowledge is applicable.

Request-based parsers did not store knowledge at as general a level as possible. Thus, a learning system using requests would not be able to tell when already-known requests would apply to new situations, and therefore knowledge would have to be re-learned for these new situations. For example, assume that a request-based parser learned a new verb. Because knowledge about verbs would be stored in the dictionary entries of each individual verb, the parser would not be able to determine which of the requests that it knew for other verbs would apply to the new verb. For example, the verb "shot" might have requests to find its subject and direct object that would apply, with only minor modifications, to a newly-learned verb. However, "shot" would also have requests that would not apply to other verbs, such as a request looking for the preposition "in" following the verb, marking the BODYPART of the victim which was wounded. Because all of these requests are simply stored in the dictionary definition of "shot," though, the parser would have no way of knowing which requests applied to the new verb. Thus, the parser would need to learn a completely new set of requests for the new verb, re-learning much of its knowledge about other verbs.

In contrast, parsing knowledge in MOPTRANS is stored at as general a level as possible. Thus, a learning system using MOPTRANS' organization of parsing knowledge would be able to apply knowledge that it learned in one situation to other, similar situations. This would avoid problems of re-learning that would be encountered in a system using request-based parsing knowledge. If a system using MOPTRANS' organization of syntactic knowledge learned a new verb, it would be able to tell which of its parsing rules applied to this verb, because parsing knowledge would be stored in terms of categories such as "verb." Thus, existing Subject and Direct Object rules would apply to the new verb, but rules like the request looking for "in" after the verb "shot" would not apply, because these rules would be stored under individual verbs.

8.2 Future Work

I have tried to motivate the theory of parsing which I have presented in this thesis in part by discussing examples which are problematic for previous natural language understanding systems. For example, I discussed examples of vague or ambiguous words which would present difficulties for syntax-based machine translation systems, and for previous conceptual analyzers. I have also presented examples of syntactically complex sentences which are difficult for request-based parsers. These problematic examples point to the shortcomings of previous theories of natural language processing, and have motivated the structure of the MOPTRANS parser. As a result, MOPTRANS is able to handle a wider range of examples than these previous systems.

In a similar way, I will now discuss some examples which are problematic for the MOPTRANS parser, indicating the shortcomings of this theory of natural language processing. These examples will point to some of the areas in which more research is necessary.

Frame Selection and Representational Issues

Before we can design a foolproof natural language system which produces the correct representation for a large class of texts, we must first know how to represent all of the different sorts of conceptual entities which the texts can be about. Thus, one of the major issues which must be investigated further is that of representation.

To tackle the problem of frame selection and word disambiguation, the MOPTRANS parser relies heavily on its hierarchically-organized system of frames. Thus, MOPTRANS' frame selection abilities are only as good as its representational system. If it is not possible to represent a distinction which must be captured in order to disambiguate a word, then the frame selection process will not succeed for that word.

Recall that in chapter 5, I discussed the frame selection process for the word "seized." "Seized" was defined as a GET-CONTROL. Depending on its OBJECT, the concept refinement rules might select a more specific frame for "seized." For example, for the sentence "A gunman seized control of a Boeing 727 and diverted it to Cuba," MOPTRANS would choose the frame HIJACK, since the prototypical OBJECT of a HIJACK was a VEHICLE.

There are stories for which MOPTRANS will incorrectly choose the frame HIJACK on the basis of its knowledge that the OBJECT of a HIJACK should be a VEHICLE. This is because its representation of vehicles does not make distinctions that need to be made to determine if the vehicle is being hijacked. Consider the following examples:

> Terrorists seized control of the space shuttle and demanded a $5 million ransom.
>
> A gunman seized control of a cable car in San Francisco today and held the riders at gunpoint.

In these examples, various properties of the seized vehicles indicate that a hijacking has not occurred, even though the OBJECT of "seized" is a VEHICLE. It would be foolish to hijack the space shuttle, since it could not take a hijacker to a useful destination. Similarly, since a cable car cannot leave its cable, it could not be hijacked.

Because the representations used in MOPTRANS do not make the distinctions necessary to determine that these vehicles cannot be hijacked, the MOPTRANS parser would select the HIJACK frame for these examples. To avoid this problem, we would need to embellish the representations of VEHICLEs with information that could be used

to determine whether or not a vehicle could be hijacked. In general, there is no easy solution to this sort of problem. For these examples, we could propose a frame like VEHICLE-WHICH-CAN-BE-HIJACKED, assign this as the OBJECT prototype of HIJACK, and only define words which refer to vehicles that can be hijacked in terms of this frame. However, this solution is quite *ad hoc,* and the number of such frames that might be required to make all the distinctions that would be needed in the domain of terrorism would be ridiculously large. Instead, it seems that the solution must involve working out a good way to represent the fact that the space shuttle travels in space, and that this is not a desirable place for hijackers to go; and that since a cable car travels on a cable, it cannot be diverted to go somewhere that it is not supposed to go. Representational problems like these must be solved before the frame selection problems which hinge on them can be attacked.

Representational problems also are responsible for difficulties which MOPTRANS has with some types of syntactic constructions. For example, in chapter 7 I presented a sentence which the Conjunction Rule would parse incorrectly:

I know John and Mary saw Fred this morning.

Although the preferable reading for this sentence is the same as "I know that John and Mary saw Fred this morning," MOPTRANS would choose the other interpretation. The reason for this is that MOPTRANS has no way of judging what conceptual elements can and cannot be conjoined. Somehow, it seems awkward to conjoin "know" and "saw," probably in part because of the different tenses of the verbs, but also probably because of the nature of the concepts underlying these verbs. The two concepts do not go well together, and thus the meaning of the sentence which does not conjoin them is preferred. It seems that the solution to problems like conjunction also must await further research on representation. Before we can write rules which determine whether or not two concepts can be conjoined, we must be able to represent the distinctions that must be made in order to make this determination.

Making the Wrong Inference

Natural language texts can be misleading. Thus, people, as well as automated natural language systems, must sometimes make the wrong inference about what a text means, requiring them to later correct their mistake.

Because texts can be misleading, the frame selection method used by the MOPTRANS parser must sometimes be misled. For example, in chapter 5 I discussed the following sentence, from (Schank, Birnbaum, and Mey, 1983):

John got a TV at Macy's.

Given the appropriate frames in MOPTRANS's conceptual hierarchy, MOPTRANS can select the frame BUY for this sentence, because of the OBJECT and LOCATION fillers of the ATRANS representing "got." However, this inference could be in error:

John got a TV at Macy's as a prize for being their millionth customer.

In this case, information later in the sentence indicates that the frame WIN is more appropriate. Thus, in this example, MOPTRANS would not choose the correct frame.

To overcome this problem, we could design frame selection rules which were more conservative. In other words, we could delay the choosing of a more specific frame until we are sure that all other frames are eliminated as possibilities. However, this approach

does not seem feasible. There are examples in which the choice of a frame would have to be delayed for a long time:

> John got a TV at Macy's. He had been wanting one for a long time, but he had no money. He took his gun to the store and pointed it at the salesman's head.

Here we see that the cue that BUY is not the correct frame does not appear until two sentences later in the story. In general, is is difficult to put a limit on how long we would have to wait to be sure that no other frame could apply. Thus, it is not practical to wait to choose a frame until all other frames are eliminated. Instead, we must accept that any frame selection process must sometimes be misled, and therefore must be able to undo its mistaken inferences.

The MOPTRANS system is not capable of abandoning a frame once the concept refinement rules have chosen it. As a result, it must choose the incorrect frame in cases where the text is misleading, such as the two examples above. How a natural language system can undo erroneous inferences is a topic for future work.

Language Learning

I have made claims that the MOPTRANS parser's organization of knowledge is more amenable to learning than previous conceptual parsers, such as request-based parsers. Thus, a topic for future research is how MOPTRANS' organization of knowledge can be applied to learning.

MOPTRANS' organization of knowledge would be applicable to a system which learned a *second* language. Such a system would start with a mastery of one language in a limited domain, and would then be taught, through natural language communication, the syntactic rules and vocabulary for a second language. MOPTRANS' organization of parsing knowledge seems amenable to this task, because of its generalized syntactic knowledge. Statements that a tutor might want to make about a second language, such as "In German, the verb comes at the end of a relative subclause, with its direct object and all prepositional phrases before it," would correspond more directly to the form of MOPTRANS' internal rules than would be true for, say, a request-based parser. Thus, it seems that the task of translating syntactic rules expressed in natural language into a form usable by the parser would be easier for a system with MOPTRANS' organization of knowledge than for a request-based system.

Appendix 1: Output of the MOPTRANS System

This appendix contains the stories parsed by the MOPTRANS parser, and the representations which the parser produces. MOPTRANS can parse versions of many of the stories in several different languages. For each story, the representation produced by the parse of the English version of the story is given. The representations produced by MOPTRANS for other languages are similar, if not identical. For the non-English stories, the computer-generated English translation is also shown.

```
Story 1:

English:
At least 60 peasants were executed by a firing squad of men wearing
olive-colored uniforms in San Pedro Perulapan.  The victims were
tried and then executed in the town plaza by guerrillas who accused them of
collaborating with the government.

Spanish:
Por lo menos 60 campesinos fueron fusilados por un grupo de guerrilleros
vestidos con trajes verde olivo en San Pedro Perulapan.

French:
Du moins 60 paysans ont ete executes par un peloton d'execution d'
des hommes portant un uniforme a San Pedro Perulapan.

German:
Mindestens 60 Bauer wurden von einer Gruppe Maenner in olivgruenen
Kleidung bei San Pedro Perulapan zurechtgestellt.

Chinese:
zai shengbideluo peirulapan, zhishao 60 ge nongmin bei shenchuan
ganlanse zhifu de xingxingdui chujue le.

Final representation:

ASS0 =
 CONCEPT ASSIST
 ACTOR   HUM0 =
           CONCEPT PERSON
           NUMBER  AT-LEAST 60
 OBJECT  ORG1 =
           CONCEPT AUTH-ORG
MTR0 =
 CONCEPT ACCUSE
 ACTOR   HUM1 =
           CONCEPT TERRORIST
           GENDER  MALE
           ORG     ORG0 =
                     CONCEPT TERRORIST-ORG
```

```
                    MEMBERS HUM1
          WEARING OBJO =
                    CONCEPT CLOTHING
                    COLOR   OLIVE-COLORED
 OBJECT  ASSO
 RECIP   HUMO
EXE1 =
 CONCEPT EXECUTE
 ACTOR   HUM1
 PLACE   LOCO =
           CONCEPT CITY
           #NAME   SAN PEDRO PERULAPAN
 TIME    LATER
 OBJECT  HUMO
ATTO =
 CONCEPT TRIAL
 PLACE   LOCO
 OBJECT  HUMO

Total time:  199490 msecs.
NIL
```

Translation:

Men from a firing squad wearing olive-colored uniforms executed at least 60 peasants in the city of San Pedro Perulapan.

Story 2:

English:
A criminal, Roger Fidel Morales Gonzalez, was killed by the patrolman who was driving him here from Tierra Azul. The convict tried to escape by jumping from the vehicle, but the patrolman fatally shot him, according to a responsible police source.

Spanish:
El reo Roger Fidel Morales Gonzalez fue matado por la patrulla que lo conducia en una camioneta desde Tierra Azul hacia esta ciudad.

French:
Un criminel, Roger Fidel Morales Gonzales, a ete tue par le policier qui le conduisait ici de la Tierra Azul.

German:
Ein Verbrecher, Roger Fidel Morales Gonzalez, wurde von dem Polizisten der ihn von Tierra Azul hierher fuhr, getoetet.

Chinese:
zuifan luojie feider molaer gongchaleci bei xunluoduiyuan dasi le.

Final representation:

```
SHOO =
```

```
 CONCEPT      SHOOT
 OBJECT       HUM3 =
                CONCEPT BAD-GUY
                GENDER  MALE
                #NAME   ROGER FIDEL MORALES GONZALEZ
 RESULT       DEAO =
                CONCEPT   DEAD
                R1        HUM3
                RESULT-OF SHOO
 ACTOR        HUM4 =
                CONCEPT AUTHORITY
 ACCORDING-TO HUM6 =
                CONCEPT PERSON
PTR19 =
 CONCEPT PTRANS
 ACTOR   HUM3
 FROM    LOC1 =
           CONCEPT PROX-PART
           R1      OBJ0 =
                     CONCEPT VEHICLE
ATT0 =
 CONCEPT ATTEMPT
 ACTOR   HUM3
 OBJECT  ESC0 =
           CONCEPT       ESCAPE
           ACTOR         HUM3
           ESC-DEEP-SUBJ HUM3
           METHOD        PTR19
PTR5 =
 CONCEPT PTRANS
 ACTOR   HUM4
 OBJECT  HUM3
 FROM    LOC0 =
           CONCEPT CITY
           #NAME   TIERRA AZUL
 TO      HERE

Total time:  193959 msecs.
NIL
```

Translation:

A patrolman who shot a convict, Roger Fidel Morales Gonzalez, to death was driving him to here from the city of Tierra Azul. The convict tried to escape.

Story 3:

English:
Presumed Basque separatist guerrillas ambushed two national police cars with explosives thursday night, wounding six policemen.

Spanish:

El jueves por la noche guerrillas Bascas emboscaron a dos vehiculos de la guardia nacional, utilizando explosivos e hiriendo a seis soldades.

French:
Des guerilleros Basques separatistes supposes ont embusque deux voitures de la police nationale avec des explosifs jeudi soir, blessant six policiers.

German:
Vermutete baskische Guerrillen fiel 2 Polizeiwagen am Donnerstag Nacht mit Sprengstoff ueber und verwundeten 6 Polizisten.

Chinese:
xingqisi yiewan, basike dulizhuyi youjidui xianyifenzi yong zhayiao fuji le er liang guomin jingche, dashang le liu ming jingcha.

Final representation:

```
HAR2 =
 CONCEPT HARM-PERSON
 ACTOR   HUM7 =
           CONCEPT     TERRORIST
           NATIONALITY LOC2 =
                         CONCEPT NATION
                         #NAME   BASQUE
           STATUS      PRESUMED-TO-BE
 OBJECT  HUM9 =
           CONCEPT AUTHORITY
           NUMBER  6
 RESULT  INJ0 =
           CONCEPT   INJURED
           R1        HUM9
           RESULT-OF HAR2
HAR1 =
 CONCEPT EXPLODE-BOMB
 INST    OBJ2 =
           CONCEPT BOMB
           INST-OF HAR1
 ACTOR   HUM7
 TIME    INS2 =
           CONCEPT     INSTANCE
           TIME-OF-DAY NIGHT
           DAY         THURSDAY
 OBJECT  OBJ1 =
           CONCEPT  VEHICLE
           OWNED-BY ORG1 =
                      CONCEPT AUTH-ORG
                      OWNS    OBJ1
           NUMBER   2

Total time:  59300 msecs.
NIL
```

Translation:

Presumed Basque guerrillas who ambushed 2 cars owned by the police with explosives on Thursday wounded 6 policemen.

Story 4:

English:
Pedro Abren Almagro, an industrialist originally from Cuba, now a resident of Guipuzcoa de Orio, has been kidnapped early this morning in Guipuzcoa Province, according to reliable sources.

Spanish:
El industrial Pedro Abren Almagro, de origen Cubano, residente en la localidad Guipuzcoana de Orio, ha sido secuestrado esta madrugada en la provincia de Guipuzcoa, indicaron fuentes competentes.

French:
Pedro Abren Almagro, industrialiste originalement de la Cuba maintenant resident de Guipuzcoa de Orio, a ete enleve tot ce matin dans la province de Guipuzcoa, selon des sources sures.

German:
Pedro Abren Almagro, ein Fabrikseigentuemer urspruenglich von Kuba und jetzt ein Einwohner von Guipuzcoa de Orio, wurde heute frueh in Guipuzcoa Provinz entfuehrt, laut vertrauter Quellen.

Chinese:
genju kekao laiyuan , yuanji guba, xianzhai shi guipuchikeya aoliao jumin de bideluo abulun amageluo jinri lingchen zai guipuchikeya sheng zaodao bangjia.

Final representation:

```
KIDO =
 CONCEPT      KIDNAP
 OBJECT       HUM10 =
                CONCEPT      PERSON
                #NAME        PEDRO ABREN ALMAGRO
                NATIONALITY  LOC3 =
                               CONCEPT NATION
                               #NAME   CUBA
                RESIDENCE    LOC4 =
                               CONCEPT CITY
                               #NAME   GUIPUZCOA DE ORIO
 SETTING      LOC5 =
                CONCEPT     LOCATION
                SETTING-OF KIDO
 TIME         INS4 =
                CONCEPT      INSTANCE
                TIME-OF-DAY  EARLY
                DAY          TODAY
```

```
 ACCORDING-TO HUM13 =
               CONCEPT PERSON
```

Total time: 87217 msecs.
NIL

Translation:

A Cuban industrialist, Pedro Abren Almagro, resident of the city of Guipuzcoa de Orio was kidnapped today in a province according to sources.

Story 5:

English:
25-year-old Rosa Areas is still in Trinity Adventist Hospital after being shot and wounded by an EPS soldier, Jose de la Cruz Quintanilla, according to members of her family.

Spanish:
Todavia se encuentra internada en el hospital Adventista de la Trinidad la joven de 25 anos Rosa Areas, la que fue herida de bala, segun el testimonio de sus familiares, por un uniformado de EPS, Jose de la Cruz Quintanilla.

French:
Rosa Areas, femme de 25 ans, reste toujours a l'holpital Trinity Adventist apres etre atteinte et blessee par un soldat de l'EPS, Jose de la Cruz Quintanilla, selon des membres de sa famille.

German:
25-jaehrige Rosa Areas ist noch in dem Spital nachdem ein Soldat, Jose de la Cruz Quintanilla, sie schoss und verwundete, laut Mitglieder ihrer Familie.

Final representation:

```
HAR3 =
 CONCEPT      HARM-PERSON
 ACTOR        HUM15 =
                CONCEPT AUTHORITY
                #NAME   JOSE DE LA CRUZ QUINTANILLA
 OBJECT       HUM14 =
                CONCEPT PERSON
                #NAME   ROSA AREAS
                AGE     YEA0 =
                          CONCEPT YEAR
                          NUMBER  25
                IS-AT   LOC6 =
                          CONCEPT HOSPITAL
 RESULT       INJ1 =
                CONCEPT    INJURED
                R1         HUM14
                RESULT-OF  HAR3
 ACCORDING-TO HUM16 =
```

```
                 CONCEPT PERSON
                 ORG     ORG3 =
                           CONCEPT ORGANIZATION
                           MEMBERS HUM16
SHO1 =
 CONCEPT SHOOT
 OBJECT  HUM14
 BEFORE  IN2 =
          CONCEPT IS-AT
          R2      LOC6
          R1      HUM14
          AFTER   SHO1

Total time:  104992 msecs.
NIL
```

Translation:

A 25-year-old woman, Rosa Areas, is at the hospital after a soldier, Jose de la Cruz Quintanilla, shot and wounded her according to members of the family.

Story 6:

English:
Police are searching for a presumed sex maniac who beat and stabbed to death a 55-year-old woman.

Spanish:
La policia realiza intensas diligencias para capturar a un presunto maniatico sexual que dio muerte a golpes y a punaladas a una mujer de 55 anos, informaron fuentes allegadas a la investigacion.

French:
La police cherche un maniac sexuel suppose qui aurait battu a mort une femme de 55 ans.

German:
Die Polizei suchen einen vermuteten Verbrecher der eine 55 -jaehrige Frau schlug und toetete.

Chinese:
jingcha zhengzai sousuo yi ge ouda bingqie cisi le yi ming 55 sui de funu de xinggongjikuang xianyifan.

Final representation:

```
HAR4 =
 CONCEPT HARM-PERSON
 ACTOR   HUM18 =
          CONCEPT BAD-GUY
          STATUS  PRESUMED-TO-BE
 OBJECT  HUM19 =
          CONCEPT PERSON
```

```
            GENDER  FEMALE
            AGE     YEA1 =
                      CONCEPT YEAR
                      NUMBER  55
  RESULT  DEA1 =
            CONCEPT   DEAD
            R1        HUM19
            RESULT-OF HAR4
FIN0 =
  CONCEPT POLICE-SEARCH
  OBJECT  HUM18
  ACTOR   HUM17 =
            CONCEPT AUTHORITY
            ORG     ORG4 =
                      CONCEPT AUTH-ORG
                      MEMBERS HUM17

Total time:  53206 msecs.
NIL
```

Translation:

The police are searching for a presumed sex maniac who beat a
55-year-old woman to death.

Story 7:

English:
Red Cross ambulances rushed two young women whose hands had been injured as
the result of a bomb to Manolo Morales hospital.

Spanish:
Ambulancias de la Cruz Roja trasladaron al hospital Manolo Morales
a dos jovencitas que sufrieron mutilaciones de sus manos a causa
de explosion de una bomba.

French:
Les ambulances de la Croix rouge ont transporte d'urgence deux jeunes
filles, dont les mains avaient ete blessees par suite d'une bombe, a
l'holpital Manolo Morales.

German:
Ein Rotkreutzkrankenwagen hastete 2 junge Frauen deren Haende von
einer Bombe verwundet wurden nach Manolo Morales Spital.

Chinese:
hongshizi jijiuche jiang zai yi ci baozha shijian zhong zhashang
le shou de er ming nianqing de funu jisu song wang mannuoluo
molaersi yiyuan.

Final representation:

```
EXP0 =
  CONCEPT EXPLODE-BOMB
```

```
 INST    OBJ6 =
           CONCEPT BOMB
           INST-OF EXPO
 OBJECT  HUM21 =
           CONCEPT PERSON
           GENDER  FEMALE
           B-PART  OBJ5 =
                     CONCEPT BODYPART
           AGE     YOUNG
           NUMBER  2
 RESULT  INJ2 =
           CONCEPT   INJURED
           R1        HUM21
           RESULT-OF EXPO
PTR99 =
 CONCEPT PTRANS-BY-AMBULANCE
 OBJECT  HUM21
 TO      LOC7 =
           CONCEPT HOSPITAL
 INST    OBJ4 =
           CONCEPT  AMBULANCE
           OWNED-BY ORG5 =
                      CONCEPT MEDICAL-ORG
                      OWNS    OBJ4
                      #NAME   RED CROSS
           INST-OF  PTR99

Total time:  80114 msecs.
NIL
```

Translation:

2 young women who were injured by a bomb in the hands were rushed by an ambulance owned by the Red Cross to the hospital.

Story 8:

English:
Three bomb attacks perpetrated last night in Marsella were attributed to the National Liberation Front of Corceja, according to an anonymous telephone call to the media.

Spanish:
Tres atentados con explosivos perpetrados antenoche en Marsella fueron atribuidos al Frente de Liberacion Nacional de Corceja por un comunicante anonimo en llamada telefonica a medios informativos.

French:
Trois attaques a bombes, perpetrees hier soir a Marsella, ont ete attribuees au Front de liberation national de Corceja, selon un coup de telephone anonyme au media.

German:
Drei Bombenangriffe gestern Nacht in Marsella wurden der National

Liberation Front of Corceja zugeschrieben, laut eines anonymen Telefonanrufes.

Chinese:
genju xinwenmeijie shoudao de niming dianhua, zuotian yiewan zai masaila fashen de san qi zhadan xiji shijian shi kesheya minzu jiefangzhenxian gan de.

Final representation:

```
HAR5 =
 CONCEPT EXPLODE-BOMB
 INST    OBJ7 =
           CONCEPT BOMB
           INST-OF HAR5
 ACTOR   HUM22 =
           CONCEPT TERRORIST
           ORG     ORG6 =
                     CONCEPT TERRORIST-ORG
                     MEMBERS HUM22
                     #NAME   NATIONAL LIBERATION FRONT OF CORCEJA
 PLACE   LOC8 =
           CONCEPT CITY
           #NAME   MARSELLA
 TIME    INS5 =
           CONCEPT     INSTANCE
           TIME-OF-DAY NIGHT
           DAY         YESTERDAY
 NUMBER  3
```

Total time: 71745 msecs.
NIL

Translation:

Terrorists from the National Liberation Front of Corceja perpetrated 3 attacks with a bomb last night in the city of Marsella.

Story 9:

English:
Members of a guerrilla group, Popular Liberation Army, killed seven people and injured five others during an assault Saturday on a ranch.

Spanish:
Un comando del grupo guerrillero Ejercito Popular de Liberacion dio muerte el Sabado a siete personas e hirio a otras cinco durante un asalto perpetrado a una hacienda.

French:
Membres d'un groupe de guerilleros, l'Armee de liberation populaire, ont tue sept personnes pendant un assaut samedi sur un ranch.

German:

Mitglieder einer Terroristenorganization, popular liberation army, toeteten 7 Personen und verwundeten 5 Anderen in einem Angriffe auf einem Bauernhof am Samstag.

Chinese:
dazhong jiefangjun youjiduiyuan zai xingqiliu xiji
muchang zhishi, dasi qih ge ren, dashang wu ge ren.

Final representation:

```
HAR8 =
 CONCEPT     HARM
 OBJECT      LOC9 =
               CONCEPT BUILDING
 TIME        INS6 =
               CONCEPT INSTANCE
               DAY     SATURDAY
 SETTING-FOR HAR7 =
               CONCEPT HARM-PERSON
               ACTOR   HUM26 =
                         CONCEPT TERRORIST
                         ORG     ORG8 =
                                   CONCEPT TERRORIST-ORG
                                   #NAME   POPULAR LIBERATION ARMY
               OBJECT  THI0 =
                         CONCEPT PERSON
                         NUMBER  5
               RESULT  INJ3 =
                         CONCEPT   INJURED
                         R1        THI0
                         RESULT-OF HAR7
               DURING  HAR8
HAR6 =
 CONCEPT HARM-PERSON
 ACTOR   HUM26
 OBJECT  HUM27 =
           CONCEPT PERSON
           NUMBER  7
 RESULT  DEA2 =
           CONCEPT   DEAD
           R1        HUM27
           RESULT-OF HAR6
ORG9 =
 CONCEPT ORGANIZATION
```

Total time: 96304 msecs.
NIL

Translation:

Guerrillas from the Popular Liberation Army killed 7 people and wounded 5 others during an assault on Saturday on a ranch.

Story 10:

English:
A Spanish industrialist, Salvador Beneitez Nieto, was kidnapped and then assassinated by suspected leftist guerrillas, according to Guatamalen police.

Spanish:
El industrial espanol Salvador Beneitez Nieto fue secuestrado y asesinado por supuestos guerrilleros izquierdistas, segun informo la policia Guatemalteca el viernes.

French:
Un industrialiste Espagnol, Salvador Beneitez Nieto, a ete enleve par des guerilleros gauchistes soupconnes, selon la police guatemalteque.

German:
Ein spanischer Industriebesitzer, Salvador Beneitez Nieto, wurde von vermuteten linksdenkenden Guerrillen entfuehrt und dann getoetet, laut der guatemalischen Polizei.

Chinese:
genju weidimala jingfang xiaoxi, xibanya gongyiejia sawaduo bennaitechi nituo bei zuopai youjidui xianyifenzi bangjia, ranhou bei shahai.

Final representation:

```
HAR9 =
 CONCEPT      HARM-PERSON
 ACTOR        HUM29 =
                CONCEPT  TERRORIST
                POLITICS LEFT-WING
 TIME         LATER
 OBJECT       HUM28 =
                CONCEPT     PERSON
                #NAME       SALVADOR BENEITEZ NIETO
                NATIONALITY LOC10 =
                              CONCEPT NATION
                              #NAME   SPAIN
 RESULT       DEA3 =
                CONCEPT   DEAD
                R1        HUM28
                RESULT-OF HAR9
 ACCORDING-TO HUM30 =
                CONCEPT AUTHORITY
                ORG     ORG10 =
                          CONCEPT AUTH-ORG
                          MEMBERS HUM30
KID1 =
 CONCEPT KIDNAP
 OBJECT  HUM28
 ACTOR   HUM29
```

Total time: 68792 msecs.
NIL

Translation:

Left-wing guerrillas kidnapped a Spanish industrialist, Salvador Beneitez Nieto. The guerrillas assassinated him later according to the police.

Story 11:

English:
Hundreds of Afghan rebels ambushed a Soviet convoy on a deserted back road, killing at least 50 Russian soldiers before escaping with armored vehicles and mortar shells, a reliable report from within Afghanistan said Thursday.

Spanish:
Cientos de rebeles Afganistanos emboscaron un convoy Sovietico en un camino desierto y mataron a cincuenta soldados Rusos antes de escapar, declaro un reporte Afgano el jueves.

French:
Des centaines de rebelles afghans ont embusque un convoi sovietique sur une ruelle deserte, tuant du moins 50 soldats russes avant de fuire avec des vehicules blindes et des obus de mortier, un rapport sur de l'Afghanistan a dit jeudi.

German:
Hunderte afganistanische Rebellen fielen ein sovietisches Geleit auf einer leeren Strasse ueber und toeteten mindestens 50 russische Soldaten bevor sie mit Panzerwagen und Mortiergranaten entflohen, laut eines zuverlaessigen Berichtes von Afghanistan am Donnerstag.

Chinese:
shubai ming afuhan fanpanzhe zai fangliang de houlu shang fuji le shulian chedui, dasi le zhishao 50 ming eguo shibing, ranhou taodun.

Final representation:

```
ESC0 =
 CONCEPT ESCAPE
 ACTOR   HUM0 =
           CONCEPT      TERRORIST
           NATIONALITY LOC0 =
                         CONCEPT NATION
                         #NAME   AFGHANISTAN
           NUMBER       HUNDREDS
 AFTER   HAR1 =
           CONCEPT HARM-PERSON
           ACTOR   HUM0
           OBJECT  HUM1 =
                     CONCEPT      AUTHORITY
                     NATIONALITY LOC3 =
```

```
                                   CONCEPT NATION
                                   #NAME   SOVIET UNION
                      NUMBER      AT-LEAST 50
            RESULT  DEA0 =
                      CONCEPT   DEAD
                      R1        HUM1
                      RESULT-OF HAR1
            BEFORE  ESC0
HAR0 =
 CONCEPT HARM
 OBJECT  GRO0 =
            CONCEPT GROUP
            MAKE-UP OBJ0 =
                      CONCEPT VEHICLE
            PART-OF LOC1 =
                      CONCEPT NATION
                      #NAME   SOVIET UNION
                      PART    GRO0
 PLACE   LOC2 =
            CONCEPT LOCATION
            STATUS  NOT-USED
 ACTOR   HUM0

Total time:  109383 msecs.
NIL
```

Translation:

Hundreds of Afghan rebels ambushed a convoy of vehicles of the Soviet Union on a deserted road and killed at least 50 Soviet soldiers.

Story 12:

English:
Black civil rights leader Vernon Jordan was ambushed and shot in the back by an unidentified sniper in a motel parking lot Thursday.

Spanish:
Vernon Jordan, el lider de los derechos civiles para los negros, fue emboscado y herido en la espalda el jueves por un francotirador en el parqueadero de un motel.

French:
Chef de droits civils des noirs, Vernon Jordan, a ete embusque et atteint au dos par un canardeur non-identifie dans le parking d'un motel jeudi.

German:
Schwarzzivilrechtsfuehrer Vernon Jordan wurde am Donnerstag in einem Motelparkiergrund von einem unidentifizierten Schuetzen in dem Ruecken geschossen.

Chinese:
xingqisi, heiren renquan lingxiu funong yuedan zai yi jia

qicheyoukeluguan de tingchechang zaodao yi ming shenfenbuming de jujishou de fuji, beibu zhongdan.

Final representation:

```
SHOO =
 CONCEPT   SHOOT
 OBJECT    HUM2 =
             CONCEPT P-LEADER
             #NAME   VERNON JORDAN
             ORG     ORGO =
                       CONCEPT GOOD-CAUSE
                       MEMBERS HUM2
             RACE    BLACK
             PART    OBJ1 =
                       CONCEPT BODYPART
                       PART-OF HUM2
 HURT-PART OBJ1
 PLACE     LOC5 =
             CONCEPT LOCATION
 TIME      INS2 =
             CONCEPT INSTANCE
             DAY     THURSDAY
 ACTOR     HUM3 =
             CONCEPT    BAD-GUY
             ARMED-WITH OBJ2 =
                          CONCEPT GUN
                          ARMING  HUM3
             TYPE       UNIDENTIFIED
HAR2 =
 CONCEPT HARM-PERSON
 OBJECT  HUM2

Total time:  59611 msecs.
NIL
```

Translation:

An armed unidentified sniper ambushed a black leader of the civil rights movement, Vernon Jordan, and shot him in the back in a parking lot on Thursday.

Story 13:

English:
Iran today said Iraqi agents killed two men and seized a number of hostages in a raid near the border with Iraq. The official Pars news agency said the Iraqis fled across the border with the unspecified number of hostages after the attack Thursday night in the town of Sar-e Po's Zahabaad.

Spanish:
Iran declaro hoy que agentes Iraqies mataron a dos hombres y capturaron a algunos rehenes en un ataque cerca de la frontera con Iraq.

French:
Iran a dit aujourd'hui que des agents iraqiens ont tue deux hommes et
se sont empare de nombre d'otages dans un raid pres de la frontiere avec
Iraq.

German:
Iran sagte heute dass irakische Agenten waehrend eines Angriffes
in der Naehe von der irakischen Grenze 2 Maenner toeteten und mehrere
Geisel nahmen.

Chinese:
yilang jintian shuo, yilake tewu xiji yilake bianjing,
dasi le er ren, zhuazou le xuduo renzhi.

Final representation:

```
HAR3 =
 CONCEPT HARM
 PLACE   LOC10 =
           CONCEPT CITY
           #NAME   SAR-E PO ZAHABAAD
 TIME    INS2 =
           CONCEPT     INSTANCE
           TIME-OF-DAY NIGHT
           DAY         THURSDAY
 BEFORE  ESC1 =
           CONCEPT ESCAPE
           ACTOR   HUM15 =
                     CONCEPT     PERSON
                     NATIONALITY LOC7 =
                                   CONCEPT NATION
                                   #NAME   IRAQ
                     CARRYING    HUM17 =
                                   CONCEPT    HOSTAGE
                                   NUMBER     A-NUMBER-OF
                                   CARRIED-BY HUM15
           TO      LOC8 =
                     CONCEPT LOCATION
           AFTER   HAR3
MTR1 =
 CONCEPT MTRANS
 ACTOR   HUM9 =
           CONCEPT   AUTHORITY
           SPOKESMAN LOC2 =
                       CONCEPT NATION
                       #NAME   IRAN
 TIME    INS0 =
           CONCEPT INSTANCE
           DAY     TODAY
 OBJECT  HAR1 =
           CONCEPT HARM-PERSON
           ACTOR   HUM10 =
                     CONCEPT     PERSON
```

```
                    NATIONALITY LOC3 =
                                  CONCEPT NATION
                                  #NAME   IRAQ
          OBJECT  HUM11 =
                    CONCEPT PERSON
                    GENDER  MALE
                    NUMBER  2
          RESULT  DEA2 =
                    CONCEPT   DEAD
                    R1        HUM11
                    RESULT-OF HAR1
 DURING  HAR2 =
          CONCEPT     HARM
          NEAR        LOC4 =
                        CONCEPT   LOCATION
                        BORDERING LOC5 =
                                    CONCEPT NATION
                                    #NAME   IRAQ
          SETTING-FOR MTR1
```

Total time: 250743 msecs.
NIL

Translation:

Iran said today that Iraqi agents killed 2 men. The agents seized a number of hostages during a raid near the border with Iraq.

Story 14:

English:
Police said yesterday that they had arrested 11 Salvadoran guerrillas who were hiding inside a church in this city.

Spanish:
La policia informo ayer haber arrestado aqui a once guerrilleros Salvadorenos que buscaron refugio en el interior de la catedral de esta ciudad.

French:
La police a dit hier qu'ils avaient arrelte 11 guerilleros salvadoriens qui se cachaient dans une eglise dans cette ville.

German:
Die Polizei sagten gestern, dass sie 11 salvadorische Guerrillen, die sich innerthalb einer Kirche in der Stadt verstochen, arrestierten.

Final representation:

```
HIDO =
 CONCEPT HIDE
 ACTOR   HUM10 =
          CONCEPT     TERRORIST
```

```
          NATIONALITY LOC10 =
                        CONCEPT NATION
                        #NAME   EL SALVADOR
          NUMBER       11
 PLACE   LOC11 =
          CONCEPT BUILDING
ARR0 =
 CONCEPT ARREST
 ACTOR   HUM9 =
          CONCEPT AUTHORITY
 PLACE   LOC12 =
          CONCEPT CITY
 OBJECT  HUM10
MTR2 =
 CONCEPT MTRANS
 ACTOR   HUM8 =
          CONCEPT AUTHORITY
          ORG     ORG1 =
                    CONCEPT AUTH-ORG
                    MEMBERS HUM8
 TIME    INS4 =
          CONCEPT INSTANCE
          DAY     YESTERDAY
 OBJECT  ARR0

Total time:  65063 msecs.
NIL
```

Translation:

The police said yesterday that they arrested 11 Salvadoran terrorists who hid in a church in the city.

Story 15:

English:
Armed separatists have seized control of Espiritu Santo Island in the South Pacific's New Hebrides and are holding two government officials hostage, the government said Saturday.

Spanish:
Separatistas armados tomaron control de la isla de Espiritu Santo en las Nueva Hebrides del pacifico del sur y mantienen a dos oficiales del gobierno como rehenes, dijo el sabado el gobierno.

French:
Des separatistes armes ont saisi l'ile Espiritu Santo dans les Nouvelles-Hebrides au Pacific du sud, et tiennent deux officiels du gouvernement pour otage, le gouvernement a dit samedi.

German:
Bewaffnete Separatisten haben die Macht in Espiritu Santo-Insel von den New Hebrideen in dem suedlichen Stillen Ozean uebernommen und halten 2 Regierungsangestellte als Geisel, sagte die Regierung

am Samstag.

Chinese:
xingqiliu zhengfu shuo, wuzhuang duilzhuyizhe zhanling le aisipiliduo santuo, zhuazhu le er ming zhengfu guanyuan zhuowei renzhi.

Final representation:

```
MTRO =
 CONCEPT MTRANS
 ACTOR   HUM3 =
           CONCEPT AUTHORITY
           ORG     ORG1 =
                     CONCEPT AUTH-ORG
                     MEMBERS HUM3
 TIME    INS0 =
           CONCEPT INSTANCE
           DAY     SATURDAY
 OBJECT  GET0 =
           CONCEPT GET-CONT
           OBJECT  LOC0 =
                     CONCEPT LOCATION
           PLACE   LOC2 =
                     CONCEPT LAND
                     #NAME   NEW HEBRIDES
                     PART-OF LOC1 =
                               CONCEPT OCEAN
                               PART    LOC2
           ACTOR   HUM0 =
                     CONCEPT    TERRORIST
                     ARMED-WITH OBJ0 =
                                  CONCEPT WEAPON
                                  ARMING  HUM0
                     CONTROL    HUM1 =
                                  CONCEPT AUTHORITY
                                  ORG     ORG0 =
                                            CONCEPT AUTH-ORG
                                            MEMBERS HUM1
                                                    HUM2 =
                                                      CONCEPT HOSTAGE
                                  NUMBER  2
                                HUM2
```

Total time: 80134 msecs.
NIL

Translation:

The government said on Saturday that armed terrorists took control of an island in the New Hebrides of the South Pacific.

Story 16:

English:

Armed separatists led by a former bulldozer driver who commands an army equipped with spears and bows and arrows seized control of Espiritu Santo Island in the South Pacific and took two government officials and 10 policemen hostage, authorities said today.

Spanish:
Las autoridades anunciaron hoy que separatistas armados
dirigidos por quien fuera el chofer de un buldozer
y quien dirige una armada equipada de lanzas, arcos y flechas
tomaron control de la isla de Espiritu Santo el el pacifico del sur
y tomaron como rehenes a dos oficiales del gobierno y diez policias.

German:
Bewaffnete Separatisten, gefuehrt von einem ehemaligen
Raupenschlepperfahrer der eine Armee ausgeruestet mit Spiessen
und Bogen und Pfeilen hat, uebernahmen Espiritu Santo Insel in
suedlichen Stillen Ozean und nahmen zwei Regierungsbeamter und
zehn Polizei als Geisel, sagten die Behoerden heute.

Final representation:

```
MTR2 =
 CONCEPT MTRANS
 ACTOR   HUM12 =
           CONCEPT AUTHORITY
 TIME    INS1 =
           CONCEPT INSTANCE
           DAY     TODAY
 OBJECT  GET2 =
           CONCEPT TAKE-HOSTAGES
           OBJECT  HUM10 =
                     CONCEPT     AUTHORITY
                   HUM11 =
                     CONCEPT HOSTAGE
           ACTOR   HUM4 =
                     CONCEPT    TERRORIST
                     ORG        ORG2 =
                                  CONCEPT TERRORIST-ORG
                                  MEMBERS HUM4
                                  LEADER  HUM6 =
                                            CONCEPT P-LEADER
                     ARMED-WITH OBJ1 =
                                  CONCEPT WEAPON
                                  ARMING  HUM4
GET1 =
 CONCEPT GET-CONT
 OBJECT  LOC3 =
           CONCEPT LOCATION
 PLACE   LOC4 =
           CONCEPT OCEAN
 ACTOR   HUM4
HUM9 =
 CONCEPT AUTHORITY
 NUMBER  10
```

```
 MEM-OF  HUM10
         HUM11
HUM8 =
 CONCEPT AUTHORITY
 ORG     ORG4 =
           CONCEPT AUTH-ORG
           MEMBERS HUM10
 NUMBER  2
 MEM-OF  HUM10
         HUM11

Total time:  172127 msecs.
NIL
```

Translation:

Authorities said today that armed terrorists from an organization led by a driver who took control of an island in the South Pacific took 2 officials from the government hostage.

Story 17:

English:
Explosions in four West Bank towns maimed two Arab mayors sympathetic to the PLO today in the worst outbreak of anti-Palestinian violence in 13 years of Israeli rule.

Spanish:
Dos alcaldes que simpatizan con la Organizacion para la Liberacion de Palestina fueron mutilados durante las explosiones que afectaron cuatro pueblos del West Bank. Esto sucedio durante el peor brote de violencia anti-Palestina en los 13 anos de ocupacion Israeli.

Chinese:
jintian, zai yiselie tongzhi de 13 nian qijian de zuiyianzhongde yi ci fanbalesitan baoluan zhong, 4 zuo xian chengshi de baozha shijian yianzhong zhashang le 2 ming tongqing plo de alabo shizhang.

Final representation:

```
UNR0 =
 CONCEPT     UNREP-ACTION
 DEGREE      WORST
 SETTING-FOR INJ0 =
               CONCEPT   INJURED
               R1        HUM13 =
                           CONCEPT        AUTHORITY
                           NATIONALITY    LOC7 =
                                            CONCEPT NATION
                                            #NAME   ARABIA
                           NUMBER         2
                           SYMPATHETIC-TO ORG5 =
                                            CONCEPT TERRORIST-ORG
```

```
                                        #NAME   PLO
                RESULT-OF EXPO =
                        CONCEPT EXPLODE-BOMB
                        PLACE   LOC6 =
                                CONCEPT CITY
                                PART-OF LOC5 =
                                        CONCEPT REGION
                                        #NAME   WEST BANK
                                        PART    LOC6
                                NUMBER  4
                        OBJECT  HUM13
                        RESULT  INJO
                TIME      INS2 =
                        CONCEPT INSTANCE
                        DAY     TODAY
                DURING    UNRO
 DURING-TIME DURO =
                CONCEPT DURATION
                TYPE    YEAR
                NUMBER  13
                DUR-OF  CON3 =
                        CONCEPT    CONTROL
                        NATION-ADJ LOC8 =
                                   CONCEPT NATION
                                   #NAME   ISRAEL
                        DUR        DURO

Total time:  97618 msecs.
NIL
```

Translation:

Explosions in 4 cities of the West Bank maimed 2 Arab mayors today during the worst violence in 13 years of Israeli rule.

Story 18:

English:
Black nationalists claimed responsibility Monday for the midnight bombings at two strategic government oil refineries that set off the worst fires in South Africa's history.

Spanish:
Nacionalistas negros se declararon responsables por las explosiones ocurridas el lunes por la noche en dos estrategicas refinerias del gobierno. Las bombas produjeron los peores fuegos que se recuerdan en Sur Africa.

German:
Schwartze Nationalisten behaupteten am Montag dass sie verantwortlich waren fuer die mitternaechtlichen Bombenangriffe bei zwei strategischen Regierungsoelraffinaderien dass 2 Maenner toeteten und die schlimmste Feuer in der Geschichte von Suedafrika verursachten.

Chinese:
xingqiyi, heiren minzuzhuyizhe xuanbu, wuyle fasheng yu zhengfu de
er jia zhanlue lianyouchang de, bingqie yinqi nanfei lishishang
zuiyianzhongde dahuo de baozha shijian shi tamen gan de.

Final representation:

```
FIR2 =
 CONCEPT     FIRE
 DEGREE      WORST
 LEAD-FROM   EXP1 =
               CONCEPT EXPLODE-BOMB
               LEAD-TO FIR2
               ACTOR   HUM14 =
                         CONCEPT TERRORIST
                         RACE    BLACK
               PLACE   LOC9 =
                         CONCEPT  BUILDING
                         OWNED-BY ORG6 =
                                    CONCEPT AUTH-ORG
                                    OWNS    LOC9
                         NUMBER   2
               TIME    INS4 =
                         CONCEPT     INSTANCE
                         TIME-OF-DAY MIDNIGHT
 DURING-TIME DUR1 =
               CONCEPT DURATION
               OF      LOC10 =
                         CONCEPT NATION
                         #NAME   SOUTH AFRICA
CLA0 =
 CONCEPT CLAIM
 OBJECT  ACT0 =
           CONCEPT ACTOR
           R1      EXP1
           R2      HUM14
           TIME    INS3 =
                     CONCEPT INSTANCE
                     DAY     MONDAY
 ACTOR   HUM14
```

Total time: 89300 msecs.
NIL

Translation:

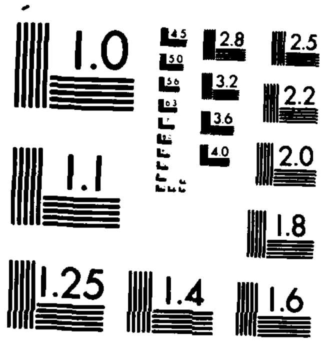
Black nationalists claimed responsibility on Monday for bombings at midnight in 2 refineries owned by the government that set off the worst fires in the history of South Africa.

Story 19:

English:
Attacks erupted on the occupied West Bank wounding at least nine

AD-A150 505 THE ORGANIZATION OF KNOWLEDGE IN A MULTI-LINGUAL INTEGRATED PARSER(U) YALE UNIV NEW HAVEN CT DEPT OF COMPUTER SCIENCE S L LYTINEN NOV 84 YALEU/CSD/RR-340 3/3

UNCLASSIFIED N00014-82-K-0149 F/G 5/7 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963 A

Palestinians including two Arab mayors seriously. A militant Jewish group headed by former American rabbi Meyer Kahane hinted that the blasts were to avenge the slayings of six Jews in Hebron.

Spanish:
Por lo menos 9 Palestinos incluyendo a dos alcaldes Arabes fueron seriamente heridos en ataques perpetrados en la zona del West Bank ocupada por Israel.

German:
In Angreife auf dem besetzten West Bank wurden mindestens neun Palestinenser einschliesslich zwei arabische Buergermeister bedenklich verletzt.

Final representation:

```
HAR6 =
 CONCEPT HARM-PERSON
 LEAD-TO RET0 =
           CONCEPT    RETALIATE
           LEAD-FROM HAR6
           METHOD     EXP0 =
                          CONCEPT EXPLODE-BOMB
 PLACE   LOC16 =
           CONCEP    ITY
           @NAME    .EBRON
 OBJECT  HUM24 =
           CONCEPT      PERSON
           NATIONALITY LOC15 =
                            CONCEPT NATION
                            @NAME   ISRAEL
           NUMBER       6
 RESULT  DEA3 =
           CONCEPT    DEAD
           R1         HUM24
           RESULT-OF HAR6
MTR2 =
 CONCEPT MTRANS
 ACTOR   HUM23 =
           CONCEPT PERSON
           ORG     ORG1 =
                     CONCEPT ORGANIZATION
                     MEMBERS HUM23
                     LEADER  HUM22 =
                               CONCEPT   P-LEADER
                               @NAME     MEYER KAHANE
                               LEADER-OF ORG1
                     MAKE-UP LOC14 =
                               CONCEPT NATION
                               @NAME   ISRAEL
                     TYPE    MILITANT
 OBJECT  EXP0
HAR4 =
 CONCEPT HARM-PERSON
```

```
PLACE   LOC11 =
          CONCEPT REGION
          #NAME   WEST BANK
OBJECT  HUM18 =
          CONCEPT   PERSON
          RELATION  LOC12 =
                      CONCEPT NATION
                      #NAME   PALESTINE
          NUMBER    AT-LEAST 9
          INCLUDING HUM19 =
                      CONCEPT     AUTHORITY
                      NATIONALITY LOC13 =
                                    CONCEPT NATION
                                    #NAME   ARABIA
                      NUMBER      2
RESULT  INJ0 =
          CONCEPT   INJURED
          R1        HUM18
          RESULT-OF HAR4
          DEGREE    SERIOUS
```

Total time: 243800 msecs.
NIL

Translation:

At least 9 Palestinians including 2 Arab mayors were wounded in the West Bank.

Story 20:

English:
Suspected black nationalist guerrillas set off satchel bombs at three strategic government refineries.

Spanish:
Presuntos guerrilleros nacionalistas negros detonaron bombas en tres estrategicas refinerias gubernamentales.

German:
Vermutete schwartz-nazionale Guerrillen entzuendeten Bomben bei drei strategischen Regierungsraffinierwerken.

Chinese:
heiren minzuzhuyi youjiduiyuan xianyifenzi zai zhengfu de 3 jia zhenlue lianyouchang jinxing baopo huodong.

Final representation:

```
LEA0 =
 CONCEPT       EXPLODE-BOMB
 INST          OBJ0 =
                 CONCEPT BOMB
                 INST-OF LEA0
```

```
 ACTOR         HUM2 =
                 CONCEPT TERRORIST
                 RACE    BLACK
 PLACE         LOC3 =
                 CONCEPT  BUILDING
                 OWNED-BY ORG0 =
                            CONCEPT AUTH-ORG
                            OWNS    LOC3
                 NUMBER   3
```

Total time: 34358 msecs.
NIL

Translation:

Black nationalists set off bombs in 3 refineries owned by the government.

Story 21:

English:
The Yugoslavian charge d'affaires and his wife and young son escaped injury early today when a bomb blast ripped through their home, the State Department said.

Spanish:
El departamento de estado anuncio que el encargado de negocios Yugoslavo junto con su esposa y su hijo escaperon ilesos de una explosion que destruyo parte de su casa hoy.

German:
Der jugoslawische Konsul und seine Frau und junge Sohn wurden nicht heute frueh verwundet als eine Bombenexplosion ihr Haus vernichtete, sagten die Behoerden.

Chinese:
guowubu shuo, jintian zaoxieshihou, dang nansilafu daiban de jiali fasheng baozha shijian shi, ta, ta de furen he nianqing de erzi xingmianyunan.

Final representation:

```
MTRO =
 CONCEPT MTRANS
 ACTOR   HUM9 =
          CONCEPT AUTHORITY
          ORG     ORG1 =
                    CONCEPT AUTH-ORG
                    MEMBERS HUM9
                    #NAME   STATE DEPARTMENT
 OBJECT  EXPO =
          CONCEPT EXPLODE-BOMB
          INST    OBJ1 =
                    CONCEPT BOMB
                    INST-OF EXPO
```

```
          PLACE   LOC5 =
                    CONCEPT BUILDING
          OBJECT  HUM8 =
                    CONCEPT     PERSON
          RESULT  INJ1 =
                    CONCEPT   INJURED
                    R1        HUM8
                    RESULT-OF EXPO
                    MODE      NEGATIVE
                    TIME      INS1 =
                                CONCEPT     INSTANCE
                                TIME-OF-DAY EARLY
                                DAY         TODAY
HUM7 =
 CONCEPT PERSON
 GENDER  MALE
 AGE     YOUNG
 MEM-OF  HUM8
HUM6 =
 CONCEPT     PERSON
 MEM-OF      HUM8
HUM5 =
 CONCEPT PERSON
 GENDER  FEMALE
 MEM-OF  HUM8
HUM4 =
 CONCEPT     PERSON
 NATIONALITY LOC4 =
               CONCEPT NATION
               #NAME   YUGOSLAVIA
 MEM-OF      HUM8

Total time:  74375 msecs.
NIL
```

Translation:

The State Department said that a bomb did not injure the Yugoslavian charges-d'affaires and his wife and his young son at home.

Story 22:

English:
Three masked gunmen late Wednesday ambushed a leading Protestant politician committed to the cause of Irish unity and shot him to death with machine guns in front of his wife and children.

Spanish:
Tres pistoleros enmascarados utilizando metralletas emboscaron y mataron en frente a su mujer y sus hijos a un lider politico protestante quien hace parte del partido de unidad Irlandesa.

German:
Drei maskierte Bewaffneten ueberfielen am Mittwoch einen

fuehrenden protestantischen Politiker der irische Einheit nachstrebt und erschossen ihn mit automatischen Gewehren vor seiner Frau und Kindern.

Final representation:

```
SHOO =
 CONCEPT SHOOT
 OBJECT  HUM11 =
          CONCEPT      PERSON
          RELIGION     PROTESTANT
          STATUS       LEADING
          COMMITTED-TO ORG2 =
                          CONCEPT GOOD-CAUSE
                          TYPE    UNITY
 RESULT  DEAO =
          CONCEPT   DEAD
          R1        HUM14
          RESULT-OF SHOO
 PLACE   LOC8 =
          CONCEPT PROX-PART
          R1      HUM17 =
                    CONCEPT     PERSON
 ACTOR   HUM10 =
          CONCEPT    PERSON
          ARMED-WITH OBJ3 =
                       CONCEPT GUN
                       ARMING  HUM10
          WEARING    OBJ2 =
                       CONCEPT CLOTHING
                       TYPE    MASK
          NUMBER     3
 INST    OBJ4 =
          CONCEPT GUN
          IN'T-OF SHOO
HAR1 =
 CONCEP'   .M-PERSON
 ACTOR   HUM10
 TIME    INS3 =
          CONCEPT INSTANCE
          DAY     WEDNESDAY
 OBJECT  HUM11
HUM16 =
 CONCEPT PERSON
 MEM-OF  HUM17
HUM15 =
 CONCEPT PERSON
 GENDER  FEMALE
 MEM-OF  HUM17
```

Total time: 120367 msecs.
NIL

Translation:

3 gunmen wearing masks ambushed a Protestant leading politician on Wednesday and shot him to death with machine guns in front of his wife and children.

Story 23:

English:
A gunman shot and wounded Reuter News Agency Middle East correspondent Bernd Debusmann in Beirut early Friday.

Spanish:
Temprano este viernes en Beirut, un hombre armado disparo e hirio a Bernd Debusman, el corresponsal de la agencia Reuter para el medio oriente.

German:
Ein Bewaffneter verwundete den Reuter Nachrichtenagenturkorrespondenten in dem Mittelosten, bernd debusman, in beirut frueh am Freitag.

Final representation:

```
HAR0 =
 CONCEPT SHOOT
 ACTOR   HUM0 =
           CONCEPT    PERSON
           ARMED-WITH OBJ0 =
                        CONCEPT GUN
                        ARMING  HUM0
 PLACE   LOC1 =
           CONCEPT CITY
           #NAME   BEIRUT
 TIME    INS1 =
           CONCEPT     INSTANCE
           TIME-OF-DAY EARLY
           DAY         FRIDAY
 OBJECT  HUM1 =
           CONCEPT      REPORTER
           #NAME        BERND DEBUSMANN
           ORG          ORG0 =
                          CONCEPT PRESS
                          MEMBERS HUM1
                          #NAME   REUTERS NEWS AGENCY
           CON-RELATION LOC0 =
                          CONCEPT NATION
                          #NAME   MIDDLE EAST
 RESULT  INJ0 =
           CONCEPT   INJURED
           R1        HUM1
           RESULT-OF HAR0
```

Total time: 46704 msecs.
NIL

Translation:

A gunman shot and wounded a reporter from Reuters News Agency, Bernd Debusmann, in the city of Beirut on Friday.

Story 24:

English:
The outlawed Irish Republican Army shot dead a part-time soldier in front of his 11-year-old son in a village store Saturday. Richard Lattimer, 30, was working in his store in the village of Newtownbutler Co Fermanagh when a gunman burst in and shot him as his son looked on. The gunman quickly fled and escaped in a waiting car.

Spanish:
La armada Irlandesa Republicana acribillo el Sabado a un soldado en frente a su hijo de 11 anos en un almacen del pueblo.

German:
Die IRA erschossen einen Teilzeitsoldaten vor seinem 11-jaehrigen Sohn am Samstag in einem Dorfsladen.

Final representation:

```
MTR3 =
 CONCEPT     SEE
 RECIP       HUM37 =
               CONCEPT PERSON
               GENDER  MALE
 INST        EYES
 SETTING-FOR SHO3 =
               CONCEPT SHOOT
               OBJECT  HUM32 =
                         CONCEPT AUTHORITY
                         GENDER  MALE
                         TYPE    PART-TIME
               RESULT  DEA5 =
                         CONCEPT   DEAD
                         R1        HUM32
                         RESULT-OF SHO3
               PLACE   LOC18 =
                         CONCEPT PROX-PART
                         R1      HUM33 =
                                   CONCEPT PERSON
                                   GENDER  MALE
                                   AGE     YEA0 =
                                             CONCEPT YEAR
                                             NUMBER  11
                                   IS-AT   LOC20 =
                                             CONCEPT BUILDING
               TIME    INS5 =
                         CONCEPT INSTANCE
                         DAY     SATURDAY
               ACTOR   HUM31 =
```

```
                    CONCEPT    TERRORIST
                    ORG        ORG4 =
                                 CONCEPT TERRORIST-ORG
                                 MEMBERS HUM31
                                 #NAME   IRISH REPUBLICAN ARMY
                                 TYPE    OUTLAWED
                    ARMED-WITH OBJ5 =
                                 CONCEPT GUN
                                 ARMING  HUM31
          DURING  MTR3
BURO =
 CONCEPT PTRANS
 ACTOR   HUM31
 DURING  UNRO =
          CONCEPT     UNREP-ACTION
          ACTOR       HUM34 =
                        CONCEPT PERSON
                        #NAME   RICHARD LATTIMER
                        AGE     YEA1 =
                                  CONCEPT YEAR
                                  NUMBER  30
          PLACE       LOC21 =
                        CONCEPT BUILDING
                        PLACE   LOC22 =
                                  CONCEPT CITY
                                  #NAME   NEWTOWNBUTLER CO FERMANAGH
          SETTING-FOR BURO
ESC3 =
 CONCEPT  ESCAPE
 ACTOR    HUM31

Total time:  341429 msecs.
NIL
```

Translation:

Terrorists from the outlawed Irish Republican Army shot a part-time soldier to death in front of his 11-year-old son in a store on Saturday.

Story 25:

English:
An armed Rumanian national whose passport apparently expired took eight people hostage today in a Queens bank and demanded the right to stay in America, police reported.

Spanish:
La policial informo un hombre armado de nacionalidad Rumana cuyo pasaporte habia expirado, tomo a 8 rehenes en un banco de Queens hoy, y demando el derecho de quedarse en los Estados Unidos.

Final representation:

```
MTRO =
```

```
 CONCEPT MTRANS
 ACTOR   HUM7 =
           CONCEPT AUTHORITY
           ORG     ORG2 =
                     CONCEPT AUTH-ORG
                     MEMBERS HUM7
 OBJECT  ALLO =
           CONCEPT ALLOW
           OBJECT  UNRO =
                     CONCEPT UNREP-ACTION
                     PLACE   LOC7 =
                               CONCEPT NATION
                               #NAME   USA
DEMO =
 CONCEPT DEMAND
 OBJECT  ALLO
 ACTOR   HUM5 =
           CONCEPT     PERSON
           NATIONALITY LOC5 =
                         CONCEPT NATION
                         #NAME   RUMANIA
           ARMED-WITH  OBJ1 =
                         CONCEPT WEAPON
                         ARMING  HUM5
GETO =
 CONCEPT TAKE-HOSTAGES
 OBJECT  HUM6 =
           CONCEPT HOSTAGE
           NUMBER  8
 PLACE   LOC6 =
           CONCEPT ORG-BUILDING
 TIME    INS3 =
           CONCEPT INSTANCE
           DAY     TODAY
 ACTOR   HUM5

Total time:  83933 msecs.
NIL
```

Translation:

The police said that an armed Rumanian person who demanded the right to stay in the USA took 8 nationals today in a bank hostage.

Story 26:

English:
Two bombs exploded today in two sections of Petah Tikva, 12 miles inland from the Tel Aviv coastline, causing no injuries or damage, police said.

German:
Zwei Bomben explodierten heute in zwei Teile von Petah Tikva, 12 Meilen inland der Kueste von Tel Aviv, verursachten aber

keine Verletzungen oder Schade, sagten die Polizei.

Final representation:

```
MTR1 =
 CONCEPT MTRANS
 ACTOR   HUM8 =
           CONCEPT AUTHORITY
           ORG     ORG3 =
                     CONCEPT AUTH-ORG
                     MEMBERS HUM8
 OBJECT  DAM1 =
           CONCEPT     DAMAGED
EXPO =
 CONCEPT EXPLODE-BOMB
 INST    OBJ2 =
           CONCEPT BOMB
           INST-OF EXPO
           NUMBER  2
 PLACE   PARO =
           CONCEPT CITY
           NUMBER  2
           PART-OF LOC8 =
                     CONCEPT CITY
                     #NAME   PETAH TIKVA
                     PART    PARO
                     REL-LOC LOC9 =
                               CONCEPT CITY
                               #NAME   TEL AVIV
 TIME    INS4 =
           CONCEPT INSTANCE
           DAY     TODAY
 RESULT  DAM1
DAMO =
 CONCEPT DAMAGED
 MEM-OF  DAM1
INJ1 =
 CONCEPT   INJURED
 RESULT-OF EXPO
 MODE      NEGATIVE
 MEM-OF    DAM1
```

Total time: 101967 msecs.
NIL

Translation:

The police said that 2 bombs exploded today in 2 sections of the city of Petah Tikva.

Story 27:

English:
Palestinian guerrillas shot and seriously wounded an Israeli border

policeman in Jerusalem and set off two bombs that exploded harmlessly near Tel Aviv Tuesday.

German:
Palestinensiche Guerrillen schossen und verwundeten einen israelischen Grenzpolizisten in Jerusalem bedenklich und entzuendeten zwei Bomben schadlos in der Naehe von Tel Aviv am Dienstag.

Final representation:

```
LEA0 =
 CONCEPT       EXPLODE-BOMB
 INST          OBJ4 =
                 CONCEPT BOMB
                 INST-OF LEA0
                 NUMBER  2
 ACTOR         HUM9 =
                 CONCEPT     TERRORIST
                 NATIONALITY LOC11 =
                               CONCEPT NATION
                               #NAME   PALESTINE
 TIME          INS5 =
                 CONCEPT INSTANCE
                 DAY     TUESDAY
 RESULT        DAM2 =
                 CONCEPT   DAMAGED
                 RESULT-OF LEA0
                 MODE      NEGATIVE
 NEAR          LOC15 =
                 CONCEPT CITY
                 #NAME   TEL AVIV
HAR1 =
 CONCEPT SHOOT
 ACTOR   HUM9
 PLACE   LOC14 =
           CONCEPT CITY
           #NAME   JERUSALEM
 OBJECT  HUM10 =
           CONCEPT     AUTHORITY
           NATIONALITY LOC12 =
                         CONCEPT NATION
                         #NAME   ISRAEL
 RESULT  INJ2 =
           CONCEPT   INJURED
           R1        HUM10
           RESULT-OF HAR1
           DEGREE    SERIOUS
```

Total time: 78425 msecs.
NIL

Translation:

Palestinian guerrillas set off 2 bombs harmlessly near the city of Tel Aviv on Tuesday and shot and wounded an Israeli policeman in the city of Jerusalem.

Story 28:

English:
A bomb planted in a locker exploded at Orly airport early today, injuring seven custodial workers and causing $250000 damage.

German:
Eine Bombe in einem Schliessschrank explodierte heute frueh bei Orly Flughafen und verwundeten sieben Reinigungsarbeiter und verursachten Schade von $250000.

Final representation:

```
EXP3 =
 CONCEPT EXPLODE-BOMB
 INST    OBJ5 =
           CONCEPT BOMB
           INST-OF EXP3
           PLACE   LOC16 =
                     CONCEPT LOCATION
 PLACE   LOC17 =
           CONCEPT LOCATION
 TIME    INS7 =
           CONCEPT     INSTANCE
           TIME-OF-DAY EARLY
           DAY         TODAY
 OBJECT  HUM11 =
           CONCEPT PERSON
           NUMBER  7
 RESULT  INJ3 =
           CONCEPT   INJURED
           R1        HUM11
           RESULT-OF EXP3
```

Total time: 63012 msecs.
NIL

Translation:

A bomb in a locker exploded in an airport today injuring 7 workers.

Story 29:

English:
A hand grenade explosion in Kabul's Soviet residential compound killed three Soviet soldiers, and a fourth was kidnapped and hacked to death, a traveler from Afghanistan said today.

German:
Eine Granatenexplosion in dem sovietischen Lager von Kabul toetete drei sovietische Soldaten, und ein Vierter wurde

entfuehrt und zum Tode gehackt, sagte ein Reisende von Afghanistan heute.

Final representation:

```
MTRO =
 CONCEPT MTRANS
 ACTOR   HUM1 =
           CONCEPT     PERSON
           NATIONALITY LOC4 =
                         CONCEPT NATION
                         #NAME   AFGHANISTAN
 TIME    INS0 =
           CONCEPT INSTANCE
           DAY     TODAY
HAR0 =
 CONCEPT HARM-PERSON
 OBJECT  THI0 =
           CONCEPT PERSON
 RESULT  DEA1 =
           CONCEPT   DEAD
           R1        THI0
           RESULT-OF HAR0
KID0 =
 CONCEPT KIDNAP
 OBJECT  THI0
EXP0 =
 CONCEPT EXPLODE-BOMB
 INST    OBJ1 =
           CONCEPT BOMB
           INST-OF EXP0
 PLACE   LOC2 =
           CONCEPT LOCATION
           PART-OF LOC1 =
                     CONCEPT NATION
                     #NAME   SOVIET UNION
                     PART    LOC2
           OF      LOC0 =
                     CONCEPT CITY
                     #NAME   KABUL
 OBJECT  HUM0 =
           CONCEPT     AUTHORITY
           NATIONALITY LOC3 =
                         CONCEPT NATION
                         #NAME   SOVIET UNION
           NUMBER      3
 RESULT  DEA0 =
           CONCEPT   DEAD
           R1        HUM0
           RESULT-OF EXP0

Total time:  81748 msecs.
NIL
```

Translation:

An Afghan traveler said today that a bomb exploded in a compound of the Soviet Union of the city of Kabul killing 3 Soviet soldiers. A fourth who was kidnapped was hacked to death.

Story 30:

English:
Leftist guerrillas ambushed a convoy of buses carrying government troops in a provincial town and 18 people died in the ensuing battle, witnesses said Thursday.

German:
Linksdenkende Guerrilien ueberfielen ein Geleit von Regierungstruppen in Autobussen in einem Provinzdorf und toeteten 18 Leute in dem nachfolgenden Kampf, sagten Zeugen am Donnerstag.

Final representation:

```
MTR1 =
 CONCEPT MTRANS
 ACTOR   HUM6 =
           CONCEPT PERSON
 TIME    INS1 =
           CONCEPT INSTANCE
           DAY     THURSDAY
 OBJECT  BAT0 =
           CONCEPT BATTLE
HAR1 =
 CONCEPT HARM
 OBJECT  GRO0 =
           CONCEPT GROUP
           MAKE-UP OBJ2 =
                     CONCEPT  VEHICLE
                     CARRYING HUM3 =
                                CONCEPT    AUTHORITY
                                ORG        ORG0 =
                                             CONCEPT AUTH-ORG
                                             MEMBERS HUM3
                                CARRIED-BY OBJ2
 PLACE   LOC6 =
           CONCEPT CITY
 ACTOR   HUM2 =
           CONCEPT  TERRORIST
           POLITICS LEFT-WING
DEA2 =
 CONCEPT DEAD
 R1      HUM4 =
           CONCEPT PERSON
           NUMBER  18
 DURING  BAT0
```

Total time: 84164 msecs.

NIL

Translation:

Witnesses said on Thursday that left-wing guerrillas ambushed a convoy of buses carrying troops from the government in a town. 18 people died during a battle.

Story 31:

English:
An unidentified gunman shot and killed a leader of Guatemala's Christian Democratic Party in a street ambush early Thursday, authorities said.

Final representation:

```
MTR2 =
 CONCEPT MTRANS
 ACTOR   HUM8 =
           CONCEPT AUTHORITY
 OBJECT  HAR3 =
           CONCEPT     HARM
           PLACE       LOC7 =
                         CONCEPT LOCATION
           TIME        INS3 =
                         CONCEPT     INSTANCE
                         TIME-OF-DAY EARLY
                         DAY         THURSDAY
           SETTING-FOR HAR2 =
                         CONCEPT SHOOT
                         ACTOR   HUM6 =
                                   CONCEPT    PERSON
                                   ARMED-WITH OBJ4 =
                                                CONCEPT GUN
                                                ARMING  HUM6
                                   TYPE       UNIDENTIFIED
                         OBJECT  HUM7 =
                                   CONCEPT   P-LEADER
                                   LEADER-OF ORG1 =
                                               CONCEPT ORGANIZATION
                                               #NAME   CHRISTIAN DEMOCRATIC PARTY
                                               LEADER  HUM7
                                               OF      LOC6 =
                                                         CONCEPT NATION
                                                         #NAME   GUATEMALA
                         RESULT  DEA3 =
                                   CONCEPT   DEAD
                                   R1        HUM7
                                   RESULT-OF HAR2
                         DURING  HAR3
```

Total time: 67463 msecs.
NIL

Story 32:

English:
Leftist guerrillas critically wounded three police guards in a daring daylight raid on the largest government office complex in San Salvador, witnesses said.

Final representation:

```
MTR3 =
 CONCEPT MTRANS
 ACTOR   HUM11 =
           CONCEPT PERSON
 OBJECT  HAR5 =
           CONCEPT     HARM
           OBJECT      LOC8 =
                         CONCEPT  LOCATION
                         OWNED-BY ORG3 =
                                    CONCEPT AUTH-ORG
                                    OWNS    LOC8
                         SIZE     LARGEST
           PLACE       LOC9 =
                         CONCEPT CITY
                         #NAME   SAN SALVADOR
           SETTING-FOR HAR4 =
                         CONCEPT HARM-PERSON
                         ACTOR   HUM9 =
                                   CONCEPT  TERRORIST
                                   POLITICS LEFT-WING
                         OBJECT  HUM10 =
                                   CONCEPT AUTHORITY
                                   ORG     ORG2 =
                                             CONCEPT AUTH-ORG
                                             MEMBERS HUM10
                                   NUMBER  3
                         RESULT  INJ0 =
                                   CONCEPT   INJURED
                                   R1        HUM10
                                   RESULT-OF HAR4
                                   DEGREE    SERIOUS
                         DURING  HAR5
```

Total time: 79043 msecs.
NIL

Story 33:

English:
Three masked gunmen who burst into the offices of a downtown bank were holding 21 hostages late Friday and threatening to kill them by a morning deadline unless a ransom was paid.

Final representation:

```
ATRO =
 CONCEPT ATRANS
 OBJECT  OBJ8 =
           CONCEPT MONEY
HAR6 =
 CONCEPT     HARM-PERSON
 ACTOR       HUM12 =
               CONCEPT    PERSON
               ARMED-WITH OBJ7 =
                            CONCEPT GUN
                            ARMING  HUM12
               WEARING    OBJ6 =
                            CONCEPT CLOTHING
                            TYPE    MASK
               NUMBER     3
               CONTROL    HUM13 =
                            CONCEPT HOSTAGE
                            NUMBER  21
 OBJECT      HUM14 =
               CONCEPT PERSON
 RESULT      DEA4 =
               CONCEPT    DEAD
               R1         HUM14
               RESULT-OF HAR6
 BEFORE-TIME INS6 =
               CONCEPT     INSTANCE
               TIME-OF-DAY MORNING
THRO =
 CONCEPT THREATEN
 OBJECT  HAR6
 ACTOR   HUM12
 UNLESS  ATRO
PTR93 =
 CONCEPT PTRANS
 ACTOR   HUM12
 TO      LOC10 =
           CONCEPT  LOCATION
           OWNED-BY LOC11 =
                      CONCEPT ORG-BUILDING
                      OWNS    LOC10

Total time:  101277 msecs.
NIL
```

Story 34:

English:
Six young men have been found machinegunned to death in two cities, victims of the extreme right-wing Squadron of Death terrorist group, authorities said Saturday.

Final representation:

```
MTR5 =
 CONCEPT MTRANS
 ACTOR   HUM18 =
           CONCEPT AUTHORITY
 TIME    INS8 =
           CONCEPT INSTANCE
           DAY     SATURDAY
 OBJECT  SHO1 =
           CONCEPT SHOOT
           OBJECT  HUM15 =
                     CONCEPT PERSON
                     GENDER  MALE
                     AGE     YOUNG
                     NUMBER  6
           RESULT  DEA5 =
                     CONCEPT   DEAD
                     R1        HUM15
                     RESULT-OF SHO1
           PLACE   LOC12 =
                     CONCEPT LOCATION
                     NUMBER  2
           ACTOR   HUM16 =
                     CONCEPT TERRORIST
                     ORG     ORG5 =
                               CONCEPT  TERRORIST-ORG
                               MEMBERS  HUM16
                               #NAME    SQUADRON OF DEATH
                               POLITICS RIGHT-WING

Total time:  96613 msecs.
NIL

Story 35:

English:
Unidentified gunmen Saturday barged into a rural church and shot to death
an Italian priest saying mass, authorities said.

Final representation:

MTR7 =
 CONCEPT MTRANS
 ACTOR   HUM21 =
           CONCEPT AUTHORITY
 OBJECT  UNR0 =
           CONCEPT UNREP-ACTION
MTR6 =
 CONCEPT MTRANS
 ACTOR   HUM20 =
           CONCEPT     PERSON
           NATIONALITY LOC14 =
                         CONCEPT NATION
                         #NAME   ITALY
 OBJECT  UNR0
```

```
SHO2 =
 CONCEPT SHOOT
 OBJECT  HUM20
 RESULT  DEA6 =
           CONCEPT   DEAD
           R1        HUM20
           RESULT-OF SHO2
 ACTOR   HUM19 =
           CONCEPT    PERSON
           ARMED-WITH OBJ9 =
                        CONCEPT GUN
                        ARMING  HUM19
           TYPE       UNIDENTIFIED
PTR144 =
 CONCEPT PTRANS
 TIME    INS9 =
           CONCEPT INSTANCE
           DAY     SATURDAY
 ACTOR   HUM19
 TO      LOC13 =
           CONCEPT BUILDING

Total time:  74654 msecs.
NIL
```

Story 36:

English:
A Yugoslav immigrant worker brandishing a double-barrelled shotgun burst into a doctor's office Monday and took 23 people hostage including three young children, police said.

Final representation:

```
MTR8 =
 CONCEPT MTRANS
 ACTOR   HUM26 =
           CONCEPT AUTHORITY
           ORG     ORG6 =
                     CONCEPT AUTH-ORG
                     MEMBERS HUM26
 OBJECT  GET0 =
           CONCEPT TAKE-HOSTAGES
           OBJECT  HUM24 =
                     CONCEPT   HOSTAGE
                     NUMBER    23
                     INCLUDING HUM25 =
                                 CONCEPT PERSON
                                 AGE     YOUNG
                                 NUMBER  3
           ACTOR   HUM22 =
                     CONCEPT     PERSON
                     NATIONALITY LOC15 =
                                   CONCEPT NATION
```

```
                              #NAME   YUGOSLAVIA
                 ARMED-WITH  OBJ10 =
                              CONCEPT GUN
                              ARMING  HUM22
PTR160 =
 CONCEPT PTRANS
 TIME    INS10 =
          CONCEPT INSTANCE
          DAY     MONDAY
 ACTOR   HUM22
 TO      LOC16 =
          CONCEPT    LOCATION
          P-OWNED-BY HUM23 =
                      CONCEPT PERSON
                      P-OWNS  LOC16

Total time:  91935 msecs.
NIL

Story 37:

English:
Police stormed a doctor's office today and shot dead a Yugoslav gunman who
had held 23 hostages for 20 hours.

Final representation:

SHOO =
 CONCEPT SHOOT
 OBJECT  HUM9 =
          CONCEPT     PERSON
          ARMED-WITH  OBJ2 =
                       CONCEPT GUN
                       ARMING  HUM9
          NATIONALITY LOC3 =
                       CONCEPT NATION
                       #NAME   YUGOSLAVIA
          CONTROL     HUM10 =
                       CONCEPT HOSTAGE
                       NUMBER  23
 RESULT  DEAO =
          CONCEPT   DEAD
          R1        HUM9
          RESULT-OF SHOO
 INST    OBJ1 =
          CONCEPT GUN
          INST-OF SHOO
 DUR     DURO =
          CONCEPT DURATION
          NUMBER  20

PTR19 =
 CONCEPT PTRANS
 ACTOR   HUM6 =
```

```
            CONCEPT AUTHORITY
            ORG     ORG1 =
                      CONCEPT AUTH-ORG
                      MEMBERS HUM6
                      MAKE-UP OBJ1
 TIME    INS1 =
           CONCEPT INSTANCE
           DAY     TODAY
 TO      LOC2 =
           CONCEPT    LOCATION
           P-OWNED-BY HUM7 =
                        CONCEPT PERSON
                        P-OWNS  LOC2
 DUR     DUR0
```

Total time: 82235 msecs.
NIL

Story 38:

English:
Three policemen abducted from their homes by left-wing terrorists were found bound and slain tuesday, the latest victims of El Salvador's political violence, authorities said.

Final representation:

```
MTR1 =
 CONCEPT MTRANS
 ACTOR   HUM15 =
           CONCEPT AUTHORITY
 OBJECT  HAR0 =
           CONCEPT HARM-PERSON
           TIME    INS2 =
                     CONCEPT INSTANCE
                     DAY     TUESDAY
           OBJECT  HUM11 =
                     CONCEPT AUTHORITY
                     NUMBER  3
           RESULT  DEA1 =
                     CONCEPT   DEAD
                     R1        HUM11
                     RESULT-OF HAR0
KID0 =
 CONCEPT KIDNAP
 ACTOR   HUM12 =
           CONCEPT TERRORIST
 OBJECT  HUM11
 FROM    LOC5 =
           CONCEPT BUILDING
```

Total time: 113195 msecs.
NIL

Story 39:

English:
Leftist guerrillas ambushed three army buses loaded with soldiers and supplies today, killing four soldiers and wounding eight others, military sources said.

Final representation:

```
MTR3 =
 CONCEPT MTRANS
 ACTOR   HUM19 =
           CONCEPT PERSON
           ORG     ORG3 =
                     CONCEPT ORGANIZATION
                     MEMBERS HUM19
 OBJECT  HAR3 =
           CONCEPT HARM-PERSON
           ACTOR   HUM17 =
                     CONCEPT    AUTHORITY
                     CARRIED-BY OBJ3 =
                                  CONCEPT  VEHICLE
                                  CARRYING HUM17
                                  OWNED-BY ORG2 =
                                             CONCEPT ORGANIZATION
                                             OWNS    OBJ3
                                  NUMBER   3
           OBJECT  THI0 =
                     CONCEPT PERSON
                     NUMBER  8
           RESULT  INJ0 =
                     CONCEPT   INJURED
                     R1        THI0
                     RESULT-OF HAR3
HAR2 =
 CONCEPT HARM-PERSON
 ACTOR   HUM17
 OBJECT  HUM18 =
           CONCEPT AUTHORITY
           NUMBER  4
 RESULT  DEA2 =
           CONCEPT   DEAD
           R1        HUM18
           RESULT-OF HAR2
HAR1 =
 CONCEPT HARM
 OBJECT  OBJ3
 ACTOR   HUM16 =
           CONCEPT  TERRORIST
           POLITICS LEFT-WING
TIM3 =
 CONCEPT TIME
 R1      CAR0 =
           CONCEPT CARRYING
```

```
          R1      OBJ3
          R2      HUM17
          TIME    INS3 =
                    CONCEPT INSTANCE
                    DAY     TODAY
 R2      INS3
```

Total time: 142404 msecs.
NIL

Story 40:

English:
Unidentified gunmen shot to death the registrar at Guatemala City's San Carlos University in a street ambush, authorities said Wednesday.

Final representation:

```
MTR1 =
 CONCEPT MTRANS
 ACTOR   HUM18 =
          CONCEPT AUTHORITY
 OBJECT  HAR0 =
          CONCEPT      HARM
          PLACE        LOC13 =
                        CONCEPT LOCATION
          SETTING-FOR SHO1 =
                        CONCEPT SHOOT
                        OBJECT  HUM12 =
                                 CONCEPT     PERSON
                                 ARMED-WITH OBJ3 =
                                             CONCEPT GUN
                                             ARMING  HUM12
                                 TYPE        UNIDENTIFIED
                        RESULT  DEA1 =
                                 CONCEPT   DEAD
                                 R1        HUM12
                                 RESULT-OF SHO1
                        PLACE   LOC12 =
                                 CONCEPT ORG-BUILDING
                        ACTOR   HUM12
                        DURING  HAR0
```

Total time: 96009 msecs.
NIL

Story 41:

English:
Iraqi security forces stormed the British embassy in Baghdad today and killed 3 gunmen who had occupied the building briefly, the state-owned Iraqi news agency said.

Final representation:

```
HAR0 =
 CONCEPT HARM-PERSON
 ACTOR   HUM0 =
           CONCEPT     PERSON
           NATIONALITY LOC1 =
                         CONCEPT NATION
                         #NAME   IRAQ
 OBJECT  HUM1 =
           CONCEPT    PERSON
           ARMED-WITH OBJ0 =
                        CONCEPT GUN
                        ARMING  HUM1
           NUMBER     3
           CONTROL    LOC5 =
                        CONCEPT BUILDING
                        PART-OF LOC2 =
                                  CONCEPT NATION
                                  #NAME   GREAT BRITAIN
                                  PART    LOC5
PTR5 =
 CONCEPT PTRANS
 ACTOR   HUM0
 TO      LOC5
 PLACE   LOC4 =
           CONCEPT CITY
           #NAME   BAGHDAD
DUR0 =
 CONCEPT DUR
 R1      CON0 =
           CONCEPT CONTROL
           R1      HUM1
           R2      LOC5
           DUR     BRIEF
 R2      BRIEF

Total time:  120884 msecs.
NIL
```

Story 42:

English:
The 21-year-old guerrilla son of a member of El Salvador's ruling junta has been captured by police after two years in hiding, authorities said Thursday.

Final representation:

```
MTR23 =
 CONCEPT MTRANS
 ACTOR   HUM147 =
           CONCEPT AUTHORITY
 TIME    INS48 =
           CONCEPT INSTANCE
```

```
          DAY      THURSDAY
 OBJECT  HID1 =
          CONCEPT HIDE
          DUR     DUR4 =
                    CONCEPT DURATION
                    TYPE    YEAR
                    NUMBER  2
                    DUR-OF  HID1
GET6 =
 CONCEPT    ARREST
 ACTOR      HUM146 =
              CONCEPT AUTHORITY
              ORG     ORG41 =
                        CONCEPT AUTH-ORG
                        MEMBERS HUM146
 OBJECT     HUM143 =
              CONCEPT TERRORIST
              GENDER  MALE
              AGE     YEA3 =
                        CONCEPT YEAR
                        NUMBER  21
              PARENT  HUM145 =
                        CONCEPT PERSON
                        ORG     ORG40 =
                                  CONCEPT ORGANIZATION
                                  MEMBERS HUM145
                                  OF      LOC97 =
                                            CONCEPT NATION
                                            #NAME   EL SALVADOR
 AFTER-TIME DUR4

Total time:  124295 msecs.
NIL
*

Story 43:

English:
The terrorists who kidnapped a Nestle Corp executive said Friday he will be
released only if the Swiss food firm comes up with an undisclosed ransom
and pays for the publication of a terrorist manifesto.

Final representation:

ATR2 =
 CONCEPT ATRANS
 ACTOR   HUM166 =
          CONCEPT PERSON
          ORG     ORG44 =
                    CONCEPT ORGANIZATION
                    MEMBERS HUM166
 OBJECT  OBJ47 =
          CONCEPT MONEY
GIV2 =
```

```
 CONCEPT GIVE-CONT
 OBJECT  HUM165 =
           CONCEPT PERSON
           GENDER  MALE
 TIME    FUTURE
 IF      ATR2
MTR26 =
 CONCEPT MTRANS
 ACTOR   HUM163 =
           CONCEPT TERRORIST
 TIME    INS52 =
           CONCEPT INSTANCE
           DAY     FRIDAY
 OBJECT  GIV2
KID7 =
 CONCEPT KIDNAP
 ACTOR   HUM163
 OBJECT  HUM164 =
           CONCEPT PERSON

Total time:  100541 msecs.
NIL

Story 44:

English:
Leftists seized three villages and assassinated 10 people in what they
claimed was retaliation for right-wing repression.

Final representation:

CON8 =
 CONCEPT     RETALIATE
 SETTING-FOR HAR33 =
               CONCEPT HARM-PERSON
               ACTOR   HUM156 =
                         CONCEPT  PERSON
                         POLITICS LEFT-WING
               OBJECT  HUM157 =
                         CONCEPT PERSON
                         NUMBER  10
               RESULT  DEA21 =
                         CONCEPT   DEAD
                         R1        HUM157
                         RESULT-OF HAR33
               DURING  CON8
CLA1 =
 CONCEPT CLAIM
 OBJECT  CON8
 ACTOR   HUM158 =
           CONCEPT PERSON
GET7 =
 CONCEPT GET-CONT
 OBJECT  LOC98 =
```

```
          CONCEPT CITY
          NUMBER  3
 ACTOR   HUM156

Total time:  69737 msecs.
NIL
```

Story 45:

Fifteen leftists armed with submachine guns Friday burst into *UPI's* office at a San Salvador radio station and tied up the news agency's correspondent and station personnel.

Final representation:

```
UNR6 =
 CONCEPT UNREP-ACTION
 ACTOR   HUM159 =
          CONCEPT    PERSON
          POLITICS   LEFT-WING
          NUMBER     15
          ARMED-WITH OBJ45 =
                      CONCEPT GUN
                      ARMING  HUM159
                      TYPE    SUBMACHINE
PTR980 =
 CONCEPT PTRANS
 ACTOR   HUM159
 TO      LOC99 =
          CONCEPT  LOCATION
          OWNED-BY ORG43 =
                    CONCEPT ORGANIZATION
                    #NAME   UNITED PRESS INTERNATIONAL
                    OWNS    LOC99
 PLACE   LOC101
TIM4 =
 CONCEPT TIME
 R1      ARM0 =
          CONCEPT ARMED-WITH
          R1      HUM159
          R2      OBJ45
          TIME    INS51 =
                   CONCEPT INSTANCE
                   DAY     FRIDAY
 R2      INS51

Total time:  106160 msecs.
NIL
```

Story 46:

English:
A right-wing terrorist group called the Squadron of Death killed 10 men Saturday including a labor leader and three others shot to death as they

ate breakfast in a restaurant, police said.

Final representation:

```
MTRO =
 CONCEPT MTRANS
 ACTOR   HUM7 =
           CONCEPT AUTHORITY
           ORG     ORG3 =
                     CONCEPT AUTH-ORG
                     MEMBERS HUM7
 OBJECT  ING1 =
           CONCEPT INGEST
           ACTOR   HUM6 =
                     CONCEPT PERSON
           PLACE   LOC0 =
                     CONCEPT LOCATION
           OBJECT  THI1 =
                     CONCEPT THING
HAR0 =
 CONCEPT SHOOT
 ACTOR   HUM0 =
           CONCEPT TERRORIST
           ORG     ORG0 =
                     CONCEPT  TERRORIST-ORG
                     MEMBERS  HUM0
                     #NAME    SQUADRON OF DEATH
                     POLITICS RIGHT-WING
 TIME    INS0 =
           CONCEPT INSTANCE
           DAY     SATURDAY
 OBJECT  HUM2 =
           CONCEPT   PERSON
           GENDER    MALE
           NUMBER    10
           INCLUDING HUM5
 RESULT  DEA0 =
           CONCEPT   DEAD
           R1        HUM2
           RESULT-OF SHO0
 DURING  ING1
```

Total time: 176444 msecs.
NIL

Story 47:

English:
Terrorists believed to be rightwing extremists shot and killed deputy state prosecutor Mario Amato Monday in a new flareup of the random political violence that has plagued Italy for 10 years.

Final representation:

```
UNRO =
 CONCEPT     VIOLENCE
 STATUS      NEW
 TYPE        POLITICAL
 SETTING-FOR SHO1 =
              CONCEPT     SHOOT
              OBJECT      HUM11 =
                            CONCEPT PERSON
                            #NAME   MARIO AMATO
              TIME        INS1 =
                            CONCEPT INSTANCE
                            DAY     MONDAY
              ACTOR       HUM8 =
                            CONCEPT  TERRORIST
                            POLITICS EXTREME
              RESULT  DEA3 =
                        CONCEPT   DEAD
                        R1        HUM11
                        RESULT-OF HAR1
              DURING      UNRO
```

Total time: 132838 msecs.
NIL

Story 48:

English:
Basque separatists bombed a hotel and a tourist development on Spain's east coast only hours before the arrival in Madrid of President Carter.

Final representation:

```
PTR73 =
 CONCEPT PTRANS
 ACTOR   HUM13 =
           CONCEPT PERSON
           TITLE   PRESIDENT
           #NAME   CARTER
 TO      LOC6 =
           CONCEPT CITY
           #NAME   MADRID
 AFTER   EXPO =
           CONCEPT EXPLODE-BOMB
           ACTOR   HUM12 =
                     CONCEPT     TERRORIST
                     NATIONALITY LOC1 =
                                   CONCEPT NATION
                                   #NAME   BASQUE
           PLACE   LOC5 =
                     CONCEPT LOCATION
           OBJECT  LOC4 =
                     CONCEPT     LOCATION
                     CONSISTS-OF (LOC3 LOC2)
           BEFORE  PTR73
```

```
Total time:  97187 msecs.
NIL

Story 49:

English:
A Basque separatist group today claimed responsibility for the killing of a
Michelin tire company executive.

Final representation:

HAR1 =
 CONCEPT HARM-PERSON
 ACTOR   HUM3 =
           CONCEPT TERRORIST
           ORG     ORG0 =
                     CONCEPT TERRORIST-ORG
                     MEMBERS HUM3
                     MAKE-UP LOC7 =
                               CONCEPT NATION
                               #NAME   BASQUE
 OBJECT  HUM5 =
           CONCEPT PERSON
 RESULT  DEA1 =
           CONCEPT   DEAD
           R1        HUM5
           RESULT-OF HAR1
CLA0 =
 CONCEPT CLAIM
 OBJECT  ACT0 =
           CONCEPT ACTOR
           R1      HAR1
           R2      HUM3
 ACTOR   HUM3

Total time:  73156 msecs.
NIL

Story 50:

English:
Masked gunmen firing submachine guns wounded one worker and kidnapped two
others in an assault on the city's troubled Coca Cola plant, officials said.

KID0 =
 CONCEPT KIDNAP
 ACTOR   HUM20 =
           CONCEPT    PERSON
           ARMED-WITH OBJ1 =
                        CONCEPT GUN
                        ARMING  HUM20
           WEARING    OBJ0 =
                        CONCEPT CLOTHING
```

```
                    TYPE    MASK
OBJECT  THI2 =
         CONCEPT PERSON
         NUMBER  2
DURING  HAR4 =
         CONCEPT     HARM
         OBJECT      LOC10 =
                       CONCEPT BUILDING
                       STATUS  TROUBLED
                       PART-OF LOC9 =
                                 CONCEPT CITY
                                 PART    LOC10
         SETTING-FOR KIDO
HAR3 =
 CONCEPT HARM-PERSON
 ACTOR   HUM20
 OBJECT  HUM21 =
         CONCEPT PERSON
 RESULT  INJ0 =
         CONCEPT   INJURED
         R1        HUM21
         RESULT-OF HAR3
 DURING  HAR4 =
         CONCEPT     HARM
         OBJECT      LOC10 =
                       CONCEPT BUILDING
                       STATUS  TROUBLED
                       PART-OF LOC9 =
                                 CONCEPT CITY
                                 PART    LOC10
         SETTING-FOR KIDO
SHO2 =
 CONCEPT SHOOT
 ACTOR   HUM20
 INST    OBJ2 =
         CONCEPT GUN
         TYPE    SUBMACHINE
         INST-OF SHO2

Total time:  151882 msecs.
NIL
```

Appendix 2: Some Detailed Examples

This appendix contains annotated program output from the MOPTRANS parser, along with the translations produced by the system's generator. This will illustrate the implementation discussed in chapter 6, as well as the parsing rules used for the individual languages which I discussed in the last chapter. The output is actually produced by MOPTRANS, with the exception of lines marked with "~".

The first example is the police investigation example which I have discussed extensively.

```
MOPTRANS created  9-Jul-84 13:14:37, ready  9-Jul-84 13:17:09

*(PARSE SP6)


Input story:

 la policia realiza intensas diligencias para capturar a un presunto
 maniatico sexual que dio muerte a golpes y a punaladas a una mujer de 55
 anos, informaron fuentes allegadas a la investigacion.

~ literally in English:  "The police are realizing intense diligent actions
~ in order to capture a presumed sexual maniac who gave death by hits
~ and by stabs to a woman of 55 years, informed sources close to the
~ investigation."

********************************************************************************
                          PARSING PROCESS BEGINS
********************************************************************************

~ The output shows the parsing rules applied, along with the state of
~ active memory resulting from the application of the rule.

~ The rule R-NEXT-WORD reads the next word in the story, and places
~ it in active memory.  R-MAKE-SYM then builds the syntactic and conceptual
~ representations for that word.

Parsing Rule applied: R-MAKE-SYM
Parsing Rule applied: R-NEXT-WORD  (la)
Parsing Rule applied: R-MAKE-SYM

********************************************************************************
ACTIVE MEMORY:

Words:                    la
Syntactic Categories:     DEF-DET
```

Conceptualizations: NIL
****

Parsing Rule applied: R-NEXT-WORD (policia)
Parsing Rule applied: R-MAKE-SYM

****
ACTIVE MEMORY:

Words: la policia
Syntactic Categories: DEF-DET N
Conceptualizations: NIL ORGO
****

Parsing Rule applied: R-HN

****
ACTIVE MEMORY:

Words: la policia
Syntactic Categories: DEF-DET HN
Conceptualizations: NIL ORGO
****

Parsing Rule applied: R-DET-1

****
ACTIVE MEMORY:

Words: policia
Syntactic Categories: NP-DEF
Conceptualizations: ORGO
****

~ The Spanish verb "realizar" (to realize), is defined in terms of the
~ conceptual relation, ACTOR. Thus, "Juan realizo la victoria" (John
~ realized the victory) builds the conceptualization (ACTOR R1 WIN R2 JOHN),
~ which is equivalent to "John is the actor of the victory."

Parsing Rule applied: R-NEXT-WORD (realiza)
Parsing Rule applied: R-TENSE
Parsing Rule applied: R-MAKE-SYM

****
ACTIVE MEMORY:

Words: policia realizar
Syntactic Categories: NP-DEF V
Conceptualizations: ORGO ACTO
****

- The subject rule places AUTH-ORG, the representation of "policia", into the R2 slot of ACTOR, because "realizar" is marked as having its subject fill this slot. Similarly, the syntactic object of "realizar" will fill the R1 slot, thus assigning the subject of "realizar" to be the ACTOR of the object of "realizar".

```
Parsing Rule applied: R-SUBJ

****************************************************************************
ACTIVE MEMORY:

Words:                     realizar
Syntactic Categories:      S
Conceptualizations:        ACTO


ACTO =
  CONCEPT ACTOR
  R2      ORGO =
            CONCEPT AUTH-ORG
****************************************************************************

Parsing Rule applied: R-NEXT-WORD  (intensas)
Parsing Rule applied: R-PLURAL
Parsing Rule applied: R-MAKE-SYM

****************************************************************************
ACTIVE MEMORY:

Words:                     realizar intensa
Syntactic Categories:      S        ADJ
Conceptualizations:        ACTO     DEGO
****************************************************************************
```

- "Diligencias" is defined as referring to the general concept, *DO*.

```
Parsing Rule applied: R-NEXT-WORD  (diligencias)
Parsing Rule applied: R-PLURAL
Parsing Rule applied: R-MAKE-SYM

****************************************************************************
ACTIVE MEMORY:

Words:                     realizar intensa diligencia
Syntactic Categories:      S        ADJ     N
Conceptualizations:        ACTO     DEGO    *DOO
****************************************************************************

Parsing Rule applied: R-HN
```

```
**************************************************************************
ACTIVE MEMORY:

Words:                    realizar intensa diligencia
Syntactic Categories:     S        ADJ     HN
Conceptualizations:       ACTO     DEGO    *DOO
**************************************************************************

Parsing Rule applied: R-ADJ-1

**************************************************************************
ACTIVE MEMORY:

Words:                    realizar diligencia
Syntactic Categories:     S        HN
Conceptualizations:       ACTO     *DOO


ACTO =                          *DOO =
  CONCEPT ACTOR                   CONCEPT *DO*
  R2       ORGO =                 DEGREE  INTENSE
             CONCEPT AUTH-ORG
**************************************************************************

Parsing Rule applied: R-NP

**************************************************************************
ACTIVE MEMORY:

Words:                    realizar diligencia
Syntactic Categories:     S        NP
Conceptualizations:       ACTO     *DOO
**************************************************************************

- At this point, when "diligencias" is assigned as the object of "realizar",
- the parser uses a Prototype Failure Rule, because it knows that when
- an ORGANIZATION, like the Police, is assigned as the ACTOR of an action,
- this actually means that some MEMBER of the ORGANIZATION performed
- the action.  Thus, instead of building (*DO* ACTOR POLICE), the
- parser builds (*DO* ACTOR POLICEMAN ORGANIZATION POLICE).

Slot-filler Specialization Demon applied (because of R2
          filler of MAKO)
Expected Filler Demon applied (because HUMO was placed in
         ?MEMBERS slot of ORGO)
Prototype Failure Rule Applied:  S-GROUP (because attempted to
          fill ACTOR slot of *DOO with ORGO)
Parsing Rule applied: R-OBJ
```

```
****************************************************************************
ACTIVE MEMORY:

Words:                    realizer diligencia
Syntactic Categories:     S        NP
Conceptualizations:       ACTO     *DOO


ACTO =                                        *DOO =
  CONCEPT ACTOR                                 CONCEPT *DO*
  R2      ORGO =                                DEGREE  INTENSE
            CONCEPT AUTH-ORG                    ACTOR   HUMO
            MEMBERS HUMO =
                      CONCEPT AUTHORITY
                      ORG     ORGO
  R1      *DOO
****************************************************************************
```

Parsing Rule applied: R-ACTOR

```
- the rule R-ACTOR is applied to verbs like "realizar", which refer to
- the relation ACTOR, after the action whose actor is being specified
- is found.  This rule assigns the action itself to be the representation
- of "realizar", so that prepositional phrases, etc., which follow the
- verb, are attached to this conceptualization, rather than to the
- relation ACTOR (e.g., in this story, the representation (*DO* ACTOR POLICE
- GOAL    GET-CONTROL) is built instead of (ACTOR R1 *DO* R2 POLICE
- GOAL    GET-CONTROL)).
```

```
****************************************************************************
ACTIVE MEMORY:

Words:                    realizer diligencia
Syntactic Categories:     S        NP
Conceptualizations:       *DOO     *DOO


*DOO =                                       *DOO =
  CONCEPT *DO*                                 CONCEPT *DO*
  DEGREE  INTENSE                              DEGREE  INTENSE
  ACTOR   HUMO =                               ACTOR   HUMO
            CONCEPT AUTHORITY
            ORG     ORGO =
                      CONCEPT AUTH-ORG
                      MEMBERS HUMO
****************************************************************************
```

Parsing Rule applied: R-NEXT-WORD (para)
Parsing Rule applied: R-MAKE-SYM

```
***********************************************************************
ACTIVE MEMORY:

Words:                    realizar diligencia para
Syntactic Categories:     S        NP         PREP
Conceptualizations:       *DOO     *DOO       PARO
***********************************************************************

Parsing Rule applied: R-NEXT-WORD  (capturar)
Parsing Rule applied: R-MAKE-SYM

***********************************************************************
ACTIVE MEMORY:

Words:                    realizar diligencia para capturar
Syntactic Categories:     S        NP         PREP INF
Conceptualizations:       *DOO     *DOO       PARO GETO
***********************************************************************

~ This is the point at which the inference process which I have described
~ at length occurs in this story, allowing the parser to infer that
~ "diligencias" refers to POLICE-INVESTIGATION.  Since *DO* is assigned
~ to be the GOAL of the GET-CONTROL, the Slot-filler Specialization Demon
~ applies, changing the representation of *DO* to FIND.  Then, since POLICE
~ is assigned as the ACTOR of the GET-CONTROL, this demon changes GET-CONTROL
~ to ARREST.  Finally, FIND is changed to POLICE-INVESTIGATION, due to
~ fact that its ACTOR is also the POLICE.

Slot-filler Specialization Demon applied (because of R1
          filler of PARO)
Slot-filler Specialization Demon applied (because of GOAL
          filler of *DOO)
Slot-filler Specialization Demon applied (because of ACTOR
          filler of GETO)
Expected Filler Demon applied (because *DOO was placed in
         ?LF slot of GETO)
Slot-filler Specialization Demon applied (because of R2
          filler of MAK1)
Prototype Failure Rule Applied:  S-GROUP (because attempted to
          fill ACTOR slot of GETO with ORGO)
Parsing Rule applied: R-PREP-INF

***********************************************************************
ACTIVE MEMORY:

Words:                    realizar capturar
Syntactic Categories:     S        S
Conceptualizations:       *DOO     GETO
```

```
*DOO =                                       GETO =
  CONCEPT POLICE-INVESTIGATION                 CONCEPT   ARREST
  DEGREE  INTENSE                              GOAL-OF   *DOO
  ACTOR   HUMO =                               ACTOR     HUMO
            CONCEPT AUTHORITY                  GOAL-OF   *DOO
            ORG     ORGO =
                      CONCEPT AUTH-ORG
                      MEMBERS HUMO
  GOAL    GETO
*****************************************************************************

Parsing Rule applied: R-NEXT-WORD  (.)
Parsing Rule applied: R-MAKE-SYM

- parsing continues, until the end of the story ...

Final representation:

MTRO =
 CONCEPT MTRANS
 ACTOR   HUM5 =
           CONCEPT PERSON
 OBJECT  HARO =
           CONCEPT HARM-PERSON
           INST    OBJ1 =
                     CONCEPT WEAPON
                     INST-OF HARO
           ACTOR   HUM2 =
                     CONCEPT BAD-GUY
                     TYPE    SEX-MANIAC
                     STATUS  PRESUMED
           OBJECT  HUM4 =
                     CONCEPT PERSON
                     GENDER  FEMALE
                     AGE     YEAO =
                               CONCEPT YEAR
                               NUMBER  55
           RESULT  DEAO =
                     CONCEPT   DEAD
                     R1        HUM4
                     RESULT-OF HARO
*DOO =
 CONCEPT POLICE-INVESTIGATION
 OBJECT  HUM2
 GOAL    GETO =
           CONCEPT   ARREST
           GOAL-OF   *DOO
           ACTOR     HUMO =
                       CONCEPT AUTHORITY
                       ORG     ORGO =
```

```
                        CONCEPT AUTH-ORG
                        MEMBERS HUMO
          OBJECT    HUM2
 ACTOR   HUMO
 DEGREE  INTENSE

Total time:  238984 msecs.
NIL
```

~ This representation is passed to the generator, which produces the
~ following translations in English and German:

Translation into English:

The police are searching for a presumed sex maniac who beat a
55-year-old woman to death.

Translation into German:

Ein Polizist suchte einen vermuteten Verbrecher. Er verwundete
eine 55 -jaehrige Frau und toetete sie.

The next story demonstrates the parser's ability to deal with multi-sentence inputs, along with its pronoun resolution abilities.

```
MOPTRANS created  9-Jul-84 14:20:39, ready 10-Jul-84 08:45:03

*(PARSE L2)

Input story:

 A criminal, Roger Fidel Morales Gonzalez, was killed by the patrolman who
 was driving him here from Tierra Azul. The convict tried to escape by
 jumping from the vehicle, but the patrolman fatally shot him, according to
 a responsible police source.

******************************************************************************
                          PARSING PROCESS BEGINS
******************************************************************************


Parsing Rule applied: R-MAKE-SYM
Parsing Rule applied: R-NEXT-WORD  (a)
Parsing Rule applied: R-MAKE-SYM

******************************************************************************
ACTIVE MEMORY:

Words:                  a
Syntactic Categories:   DET
```

```
Conceptualizations:     NIL
******************************************************************************

Parsing Rule applied: R-NEXT-WORD  (criminal)
Parsing Rule applied: R-MAKE-SYM

******************************************************************************
ACTIVE MEMORY:

Words:                  a   criminal
Syntactic Categories:   DET N
Conceptualizations:     NIL HUMO


HUMO =
  CONCEPT BAD-GUY
******************************************************************************

Parsing Rule applied: R-NEXT-WORD  (",")
Parsing Rule applied: R-HN

******************************************************************************
ACTIVE MEMORY:

Words:                  a   criminal ","
Syntactic Categories:   DET HN       NIL
Conceptualizations:     NIL HUMO     NIL
******************************************************************************

Parsing Rule applied: R-DET

******************************************************************************
ACTIVE MEMORY:

Words:                  criminal ","
Syntactic Categories:   NP-DET   NIL
Conceptualizations:     HUMO     NIL
******************************************************************************

Parsing Rule applied: R-MAKE-SYM

******************************************************************************
ACTIVE MEMORY:

Words:                  criminal ","
Syntactic Categories:   NP-DET   PUNC
Conceptualizations:     HUMO     NIL


HUMO =
```

```
  CONCEPT BAD-GUY
********************************************************************************

Parsing Rule applied: R-NP-PUNC

********************************************************************************
ACTIVE MEMORY:

Words:                   criminal
Syntactic Categories:    NP-PUNC
Conceptualizations:      HUMO
********************************************************************************

Parsing Rule applied: R-NEXT-WORD  (roger)
Parsing Rule applied: R-MAKE-SYM

********************************************************************************
ACTIVE MEMORY:

Words:                   criminal roger
Syntactic Categories:    NP-PUNC  WORD
Conceptualizations:      HUMO     NIL


HUMO =
  CONCEPT BAD-GUY
********************************************************************************

- The parser has rules telling it how to deal with undefined words.
- "Roger" is an undefined word.  Often undefined words are simply
- discarded; however, when they appear in certain positions after
- a noun phrase which refers to a person or some other object which
- can possess a name, they are assumed to be names.  This is what
- happens with "Roger Fidel Morales Gonzalez".

Parsing Rule applied: R-UND-NAME-2

********************************************************************************
ACTIVE MEMORY:

Words:                   criminal roger
Syntactic Categories:    NP-PUNC  NAME
Conceptualizations:      HUMO     NIL
********************************************************************************

Parsing Rule applied: R-NAME-AFTER-N

********************************************************************************
ACTIVE MEMORY:
```

```
Words:                    criminal roger
Syntactic Categories:     NP-PUNC  NAME
Conceptualizations:       HUMO     NIL


HUMO =
  CONCEPT BAD-GUY
  #NAME   (roger)
*****************************************************************************

Parsing Rule applied: R-NEXT-WORD  (fidel)
Parsing Rule applied: R-MAKE-SYM

*****************************************************************************
ACTIVE MEMORY:

Words:                    criminal roger fidel
Syntactic Categories:     NP-PUNC  NAME  WORD
Conceptualizations:       HUMO     NIL   NIL


HUMO =
  CONCEPT BAD-GUY
  #NAME   (roger)
*****************************************************************************

Parsing Rule applied: R-UND-NAME-3

*****************************************************************************
ACTIVE MEMORY:

Words:                    criminal roger fidel
Syntactic Categories:     NP-PUNC  NAME  NAME
Conceptualizations:       HUMO     NIL   NIL
*****************************************************************************

Parsing Rule applied: R-NAME-AFTER-N-REST

*****************************************************************************
ACTIVE MEMORY:

Words:                    criminal roger
Syntactic Categories:     NP-PUNC  NAME
Conceptualizations:       HUMO     NIL
*****************************************************************************

~ skipping on ...

~ the referent of the pronoun "him" is determined when it is assigned
~ as the OBJECT of "driving".  This occurs after "him" is made into
```

```
- an NP.

Parsing Rule applied: R-NEXT-WORD  (him)
Parsing Rule applied: R-MAKE-SYM

******************************************************************************
ACTIVE MEMORY:

Words:                    killed patrolman driving him
Syntactic Categories:     S      NP-DET    S       N
Conceptualizations:       HAR0   HUM1      PTR5    HUM2


HAR0 =                                                  HUM1 =
  CONCEPT HARM-PERSON                                     CONCEPT AUTHORITY
  OBJECT  HUM0 =
            CONCEPT BAD-GUY
            #NAME   (roger fidel morales gonzalez)
  RESULT  DEA0 =
            CONCEPT   DEAD
            R1        HUM0
            RESULT-OF HAR0
  ACTOR   HUM1

PTR5 =            HUM2 =
  CONCEPT PTRANS    CONCEPT PERSON
  ACTOR   HUM1      GENDER  MALE
******************************************************************************

Parsing Rule applied: R-NEXT-WORD  (here)
Parsing Rule applied: R-HN

******************************************************************************
ACTIVE MEMORY:

Words:                    killed patrolman driving him  here
Syntactic Categories:     S      NP-DET    S       HN   NIL
Conceptualizations:       HAR0   HUM1      PTR5    HUM2 NIL
******************************************************************************

Parsing Rule applied: R-NP

******************************************************************************
ACTIVE MEMORY:

Words:                    killed patrolman driving him  here
Syntactic Categories:     S      NP-DET    S       NP   NIL
Conceptualizations:       HAR0   HUM1      PTR5    HUM2 NIL
******************************************************************************
```

Parsing Rule applied: R-OBJ

Now that the Object Rule is applied, (PERSON GENDER MALE) is assigned as the OBJECT of the PTRANS. Then, the Object Rule attempts to find a referent. It knows that the SUBJECT of "driving" cannot be the referent, since "him" is not reflexive. This leaves only one other possibility: the criminal. Therefore, HUMO (the criminal) is chosen as the referent of the pronoun.

```
Found referent for pronoun "him":
HUMO =
 CONCEPT BAD-GUY
 #NAME   roger fidel morales gonzalez

*************************************************************************.******
ACTIVE MEMORY:

Words:                      killed patrolman driving him  here
Syntactic Categories:       S      NP-DET    S       NP   NIL
Conceptualizations:         HARO   HUM1      PTR5    HUMO NIL


HARO =                          HUM1 =                PTR5 =
  CONCEPT HARM-PERSON             CONCEPT AUTHORITY     CONCEPT PTRANS
  OBJECT  HUMO                                          ACTOR   HUM1
  RESULT  DEAO =                                        OBJECT  HUMO
            CONCEPT    DEAD
            R1         HUMO
            RESULT-OF  HARO
  ACTOR   HUM1

HUMO =
  CONCEPT BAD-GUY
  #NAME   (roger fidel morales gonzalez)
  GENDER  MALE
********************************************************************************
```

skipping on further...

Here are the contents of active memory at the end of processing the first sentence. Next, since the end of the sentence has been encountered, the parser clears active memory, to begin processing the next sentence.

In addition to active memory, the MOPTRANS parser uses a second memory, which contains all of the conceptual representations built during the story so far. This memory is referred to by pronoun resolution rules, and to determine when multiple references have been made to the same event. In the second sentence of this example, we will see some instances of this.

```
*******************************************************************************
ACTIVE MEMORY:

Words:                   killed patrolman driving tierra-azul "."
Syntactic Categories:    S      NP-DET    S       NP          NIL
Conceptualizations:      HARO   HUM1      PTR5    LOCO        NIL


HARO =                                                  HUM1 =
  CONCEPT HARM-PERSON                                     CONCEPT AUTHORITY
  OBJECT  HUMO =
            CONCEPT BAD-GUY
            #NAME   (roger fidel morales gonzalez)
            GENDER  MALE
  RESULT  DEAO =
            CONCEPT   DEAD
            R1        HUMO
            RESULT-OF HARO
  ACTOR   HUM1

PTR5 =                LOCO =
  CONCEPT PTRANS        CONCEPT CITY
  ACTOR   HUM1          #NAME   (TIERRA AZUL)
  OBJECT  HUMO
  TO      HERE
  FROM    LOCO
*******************************************************************************

Parsing Rule applied: R-PERIOD

*******************************************************************************
ACTIVE MEMORY:

Words:                   NIL
Syntactic Categories:    NIL
Conceptualizations:      NIL
*******************************************************************************

Parsing Rule applied: R-NEXT-WORD  (the)
Parsing Rule applied: R-MAKE-SYM

*******************************************************************************
ACTIVE MEMORY:

Words:                   the
Syntactic Categories:    DET
Conceptualizations:      NIL
*******************************************************************************

Parsing Rule applied: R-NEXT-WORD  (convict)
```

Parsing Rule applied: R-MAKE-SYM

```
******************************************************************************
ACTIVE MEMORY:

Words:                     the convict
Syntactic Categories:      DET N
Conceptualizations:        NIL HUM3


HUM3 =
  CONCEPT BAD-GUY
******************************************************************************
```

Parsing Rule applied: R-NEXT-WORD (tried)
Parsing Rule applied: R-HN

```
******************************************************************************
ACTIVE MEMORY:

Words:                     the convict tried
Syntactic Categories:      DET HN      NIL
Conceptualizations:        NIL HUM3    NIL
******************************************************************************
```

Parsing Rule applied: R-DET

```
******************************************************************************
ACTIVE MEMORY:

Words:                     convict tried
Syntactic Categories:      NP-DET  NIL
Conceptualizations:        HUM3    NIL
******************************************************************************
```

Parsing Rule applied: R-DEF-NP

- The parsing rule R-DEF-NP is the one which is responsible for checking
- conceptual memory to try to resolve the reference of a definite noun
- phrase (i.e., one in which a definite article is used). This rule
- examines all of the representations in conceptual memory to see if
- any match the description provided by "the convict". If exactly
- one representation matches this description, then it is the referent
- of the noun phrase. In this case, "the convict" matches the
- representation for "a criminal, Roger Fidel Morales Gonzalez,"
- in the previous sentence.

```
******************************************************************************
ACTIVE MEMORY:
```

```
Words:                    convict tried
Syntactic Categories:     NP-DET  NIL
Conceptualizations:       HUM3    NIL


HUM3 =
  CONCEPT BAD-GUY
  #NAME   (roger fidel morales gonzalez)
  GENDER  MALE
*****************************************************************************

- The verb "tried" is defined as ambiguous in MOPTRANS.  It can either
- mean ATTEMPT or TRIAL.  Thus, when the word is first encountered,
- a "dummy" representation, as discussed in chapter 6, is built.
- This representation is called ATTEMPT-OR-TRIAL.  Depending on the
- way in which the OBJECT of this dummy representation is filled in,
- the Slot-filler Specialization Demon chooses one of these meanings.

Parsing Rule applied: R-MAKE-SYM

*****************************************************************************
ACTIVE MEMORY:

Words:                    convict tried
Syntactic Categories:     NP-DET  (OR V VPP)
Conceptualizations:       HUM3    ATTO


HUM3 =                                      ATTO =
  CONCEPT BAD-GUY                             CONCEPT ATTEMPT-OR-TRIAL
  #NAME   (roger fidel morales gonzalez)
  GENDER  MALE
*****************************************************************************

Parsing Rule applied: R-SUBJ

*****************************************************************************
ACTIVE MEMORY:

Words:                    tried
Syntactic Categories:     S
Conceptualizations:       ATTO


ATTO =
  CONCEPT ATTEMPT-OR-TRIAL
  ACTOR   HUM3 =
            CONCEPT BAD-GUY
            #NAME   (roger fidel morales gonzalez)
```

```
          GENDER  MALE
******************************************************************************

- skipping on...

Parsing Rule applied: R-NEXT-WORD  (escape)
Parsing Rule applied: R-MAKE-SYM

******************************************************************************
ACTIVE MEMORY:

Words:                   tried to   escape
Syntactic Categories:    S     PREP VINF
Conceptualizations:      ATTO  #TOO ESCO


ATTO =                                              #TOO =
  CONCEPT ATTEMPT-OR-TRIAL                            CONCEPT #TO
  ACTOR   HUM3 =
          CONCEPT BAD-GUY
          #NAME   (roger fidel morales gonzalez)
          GENDER  MALE

ESCO =
  CONCEPT ESCAPE-OR-MODE
******************************************************************************

- "to escape" is assigned to be an infinitive.  Then, the rule R-TRY-TO-INF
- assigns the infinitive as the OBJECT of "tried".  At this point, since
- the OBJECT of ATTEMPT-OR-TRIAL is an ACTION, the meaning ATTEMPT is
- chosen by the Slot-filler Specialization Demon.

Parsing Rule applied: R-VINF

******************************************************************************
ACTIVE MEMORY:

Words:                   tried escape
Syntactic Categories:    S     INF
Conceptualizations:      ATTO  ESCO
******************************************************************************

Slot-filler Specialization Demon applied (because of OBJECT
          filler of ATTO)
Expected Filler Demon applied (because ESCO was placed in
         ?OBJECT slot of ATTO)
Parsing Rule applied: R-TRIED-INF

******************************************************************************
ACTIVE MEMORY:
```

```
Words:                    tried escape
Syntactic Categories:     S     S
Conceptualizations:       ATTO  ESCO


ATTO =                                                   ESCO =
  CONCEPT ATTEMPT                                          CONCEPT       ESCAPE
  ACTOR   HUM3 =                                           ACTOR         HUM3
            CONCEPT BAD-GUY
            #NAME   (roger fidel morales gonzalez)
            GENDER  MALE
  OBJECT  ESCO
******************************************************************************

- skipping on ...

******************************************************************************
ACTIVE MEMORY:

Words:                    tried escape jumping vehicle but  patrolman
                          shot
Syntactic Categories:     S     S      S       NP-PUNC CONJ NP-DET
                          (OR V VPP)
Conceptualizations:       ATTO  ESCO   PTR19   OBJO    NIL  HUM4
                          SHOO


ATTO =                                                   ESCO =
  CONCEPT ATTEMPT                                          CONCEPT       ESCAPE
  ACTOR   HUM3 =                                           ACTOR         HUM3
            CONCEPT BAD-GUY                                ESC-DEEP-SUBJ HUM3
            #NAME   (roger fidel morales gonzalez)         METHOD        PTR19
            GENDER  MALE
  OBJECT  ESCO

PTR19 =                         OBJO =             HUM4 =
  CONCEPT PTRANS                  CONCEPT VEHICLE    CONCEPT AUTHORITY
  ACTOR   HUM3
  FROM    LOC1 =
            CONCEPT PROX-PART
            R1      OBJO

SHOO =
  CONCEPT SHOOT
  RESULT  DEA1 =
            CONCEPT   DEAD
            RESULT-OF SHOO
******************************************************************************
```

Parsing Rule applied: R-SUBJ

```
****************************************************************************
ACTIVE MEMORY:

Words:                    tried escape jumping vehicle but  shot
Syntactic Categories:     S     S      S        NP-PUNC CONJ S
Conceptualizations:       ATTO  ESCO   PTR19    OBJO    NIL  SHOO


ATTO =                                                ESCO =
  CONCEPT ATTEMPT                                       CONCEPT       ESCAPE
  ACTOR   HUM3 =                                        ACTOR         HUM3
            CONCEPT BAD-GUY                             ESC-DEEP-SUBJ HUM3
            #NAME   (roger fidel morales gonzalez)      METHOD        PTR19
            GENDER  MALE
  OBJECT  ESCO

PTR19 =                         OBJO =              SHOO =
  CONCEPT PTRANS                  CONCEPT VEHICLE     CONCEPT SHOOT
  ACTOR   HUM3                                        RESULT  DEA1 =
  FROM    LOC1 =                                                CONCEPT   DEAD
            CONCEPT PROX-PART                                   RESULT-OF SHOO
            R1      OBJO                              ACTOR   HUM4 =
                                                                CONCEPT
                                                                     AUTHORITY
****************************************************************************
```

- At this point, another rule which tries to resolve references is applied.
- This rule, R-S, is called on whenever a new action is built. Just
- as with R-NP-DEF, conceptual memory is checked to see if any actions
- built earlier in the parse match the representation built for the current
- verb. In this case, (SHOOT ACTOR AUTHORITY RESULT DEAD) matches
- with (HARM-PERSON ACTOR AUTHORITY OBJECT (BAD-GUY NAME (roger fidel
- morales gonzalez)) RESULT DEAD), which was built in the first sentence
- to represent "killed". Thus, the OBJECT of the SHOOTING is filled in.

Parsing Rule applied: R-S

```
****************************************************************************
ACTIVE MEMORY:

Words:                    tried escape jumping vehicle but  shot
Syntactic Categories:     S     S      S        NP-PUNC CONJ S
Conceptualizations:       ATTO  ESCO   PTR19    OBJO    NIL  SHOO


ATTO =                                                ESCO =
  CONCEPT ATTEMPT                                       CONCEPT       ESCAPE
  ACTOR   HUM3 =                                        ACTOR         HUM3
```

```
          CONCEPT BAD-GUY                              ESC-DEEP-SUBJ HUM3
          #NAME   (roger fidel morales gonzalez)       METHOD        PTR19
          GENDER  MALE
  OBJECT  ESCO

PTR19 =                       OBJ0 =              SHO0 =
  CONCEPT PTRANS                CONCEPT VEHICLE     CONCEPT SHOOT
  ACTOR   HUM3                                      OBJECT  HUM3
  FROM    LOC1 =                                    RESULT  DEA0 =
          CONCEPT PROX-PART                                   CONCEPT   DEAD
          R1      OBJ0                                        R1        HUM3
                                                              RESULT-OF SHO0
                                                    ACTOR   HUM4 =
                                                              CONCEPT
                                                                   AUTHORITY
*******************************************************************************

Parsing Rule applied: R-NEXT-WORD  (him)
Parsing Rule applied: R-MAKE-SYM

*******************************************************************************
ACTIVE MEMORY:

Words:                   tried escape jumping vehicle but  shot him
Syntactic Categories:    S     S      S       NP-PUNC CONJ S    N
Conceptualizations:      ATT0  ESC0   PTR19   OBJ0    NIL  SHO0 HUM5


ATT0 =                                                ESC0 =
  CONCEPT ATTEMPT                                       CONCEPT       ESCAPE
  ACTOR   HUM3 =                                        ACTOR         HUM3
          CONCEPT BAD-GUY                               ESC-DEEP-SUBJ HUM3
          #NAME   (roger fidel morales gonzalez)        METHOD        PTR19
          GENDER  MALE
  OBJECT  ESC0

PTR19 =                       OBJ0 =              SHO0 =
  CONCEPT PTRANS                CONCEPT VEHICLE     CONCEPT SHOOT
  ACTOR   HUM3                                      OBJECT  HUM3
  FROM    LOC1 =                                    RESULT  DEA0 =
          CONCEPT PROX-PART                                   CONCEPT   DEAD
          R1      OBJ0                                        R1        HUM3
                                                              RESULT-OF SHO0
                                                    ACTOR   HUM4 =
                                                              CONCEPT
                                                                   AUTHORITY

HUM5 =
  CONCEPT PERSON
  GENDER  MALE
```

```
******************************************************************************

Parsing Rule applied: R-NEXT-WORD  (",")
Parsing Rule applied: R-HN

******************************************************************************
ACTIVE MEMORY:

Words:                    tried escape jumping vehicle but  shot him  ","
Syntactic Categories:     S     S      S        NP-PUNC CONJ S    HN   NIL
Conceptualizations:       ATTO  ESCO   PTR19    OBJO    NIL  SHOO HUM5 NIL
******************************************************************************

Parsing Rule applied: R-NP

******************************************************************************
ACTIVE MEMORY:

Words:                    tried escape jumping vehicle but  shot him  ","
Syntactic Categories:     S     S      S        NP-PUNC CONJ S    NP   NIL
Conceptualizations:       ATTO  ESCO   PTR19    OBJO    NIL  SHOO HUM5 NIL
******************************************************************************

- At this point, the rule R-OBJ is executed, assigning "him" as the OBJECT
- of SHOOT.  Since the representation of SHOOT was already merged with
- the representation HARM-PERSON produced by the first sentence, filling
- in the OBJECT of the SHOOTing, this slot-filling resolves the reference
- of "him".

Found referent for pronoun "him":
HUM3 =
 CONCEPT BAD-GUY
 GENDER  MALE
 #NAME   roger fidel morales gonzalez
Parsing Rule applied: R-OBJ

******************************************************************************
ACTIVE MEMORY:

Words:                    tried escape jumping vehicle but  shot him  ","
Syntactic Categories:     S     S      S        NP-PUNC CONJ S    NP   PUNC
Conceptualizations:       ATTO  ESCO   PTR19    OBJO    NIL  SHOO HUM5 NIL


ATTO =                                                 ESCO =
  CONCEPT ATTEMPT                                        CONCEPT         ESCAPE
  ACTOR   HUM3 =                                         ACTOR           HUM3
            CONCEPT BAD-GUY                              ESC-DEEP-SUBJ   HUM3
            #NAME   (roger fidel morales gonzalez)       METHOD          PTR19
            GENDER  MALE
```

```
 OBJECT  ESC0

PTR19 =                         OBJ0 =               SHO0 =
 CONCEPT PTRANS                  CONCEPT VEHICLE       CONCEPT SHOOT
 ACTOR   HUM3                                          OBJECT  HUM3
 FROM    LOC1 =                                        RESULT  DEA0 =
           CONCEPT PROX-PART                                     CONCEPT   DEAD
           R1      OBJ0                                          R1        HUM3
                                                                 RES[illegible]-OF SHO0
                                                       ACTOR   HUM4 =
                                                                 CONCEPT
                                                                      AUTHORITY
********************************************************************************
```

- *The parsing proceeds, until the end of the sentence:*

Final representation:

```
SHO0 =
 CONCEPT      SHOOT
 OBJECT       HUM3 =
                CONCEPT BAD-GUY
                GENDER  MALE
                #NAME   roger fidel morales gonzalez
 RESULT       DEA0 =
                CONCEPT   DEAD
                R1        HUM3
                RESULT-OF SHO0
 ACTOR        HUM4 =
                CONCEPT AUTHORITY
 ACCORDING-TO HUM6 =
                CONCEPT PERSON
PTR19 =
 CONCEPT PTRANS
 ACTOR   HUM3
 FROM    LOC1 =
           CONCEPT PROX-PART
           R1      OBJ0 =
                     CONCEPT VEHICLE
ATT0 =
 CONCEPT ATTEMPT
 ACTOR   HUM3
 OBJECT  ESC0 =
           CONCEPT       ESCAPE
           ACTOR         HUM3
           ESC-DEEP-SUBJ HUM3
           METHOD        PTR19
PTR5 =
 CONCEPT PTRANS
 ACTOR   HUM4
```

```
 OBJECT  HUM3
 FROM    LOCO =
           CONCEPT CITY
           #NAME   TIERRA AZUL
 TO      HERE

Total time:  193959 msecs.
NIL
*
```

The following example demonstrates the parser's abilities in German:

```
MOPTRANS created 12-Jul-84 11:49:09, ready 12-Jul-84 11:51:42

*(PARSE G13)


Input story:

 Iran sagte heute dass irakische Agenten waehrend eines Angriffes in der
 Naehe von der irakischen Grenze 2 Maenner toeteten und mehrere Geisel
 nahmen.

~ literally in English:  "Iran said today that iraqi agents during
~ a raid near the iraqi border 2 men killed and a number of hostages
~ seized."  Or, in good English: "Iran said today that iraqi agents
~ killed 2 men and seized a number of hostages in a raid near the
~ iraqi border."

********************************************************************************
                        PARSING PROCESS BEGINS
********************************************************************************


Parsing Rule applied: R-MAKE-SYM
Parsing Rule applied: R-NEXT-WORD  (Iran)
Parsing Rule applied: R-MAKE-SYM

********************************************************************************
ACTIVE MEMORY:

Words:                      Iran
Syntactic Categories:       N
Conceptualizations:         LOCO
********************************************************************************

Parsing Rule applied: R-NP

********************************************************************************
```

```
ACTIVE MEMORY:

Words:                    Iran
Syntactic Categories:     NP
Conceptualizations:       LOCO
******************************************************************************

Parsing Rule applied: R-NG

******************************************************************************
ACTIVE MEMORY:

Words:                    Iran
Syntactic Categories:     NG
Conceptualizations:       LOCO


LOCO =
  CONCEPT NATION
******************************************************************************

Parsing Rule applied: R-NEXT-WORD  (sagte)
Parsing Rule applied: R-MORPH
Parsing Rule applied: R-MAKE-SYM

******************************************************************************
ACTIVE MEMORY:

Words:                    Iran sagen
Syntactic Categories:     NG   V
Conceptualizations:       LOCO MTRO
******************************************************************************

- A Prototype Failure Rule applies here.  When a newspaper story says
- "Iran said ...", it really means "a spokesman from Iran said ...",
- since countries cannot be the ACTORs of an MTRANS.  Thus, a rule
- applies which builds the representation (MTRANS ACTOR (PERSON SPOKESMAN
- IRAN)).

Prototype Failure Rule Applied:  S-MTRANS-NATION (because attempted to
          fill ACTOR slot of MTRO with LOCO)
Parsing Rule applied: R-NG-V-CLASS

******************************************************************************
ACTIVE MEMORY:

Words:                    sagen
Syntactic Categories:     V
Conceptualizations:       MTRO
```

```
MTRO =
  CONCEPT MTRANS
  ACTOR   HUM0 =
          CONCEPT   PERSON
          SPOKESMAN LOC0 =
                    CONCEPT NATION
******************************************************************************
```

- The parse continues until the clause, beginning with "dass irakische
- Agenten" (that Iraqi agents).

```
******************************************************************************
ACTIVE MEMORY:

Words:                      sagen dass Agente
Syntactic Categories:       V     CLM  NG
Conceptualizations:         MTRO  NIL  HUM1


MTRO =                                      HUM1 =
  CONCEPT MTRANS                              CONCEPT     PERSON
  ACTOR   HUM0 =                              NATIONALITY LOC1 =
          CONCEPT   PERSON                                CONCEPT NATION
          SPOKESMAN LOC0 =                                #NAME   (iraq)
                    CONCEPT NATION
  TIME    INS0 =
          CONCEPT INSTANCE
          DAY     TODAY
******************************************************************************
```

- At this point, a "dummy" representation is built for the clause, and
- stored under the word "dass". This is so that PP's, etc., can be
- attached to this representation, to facilitate the ability to infer
- the action before the verb is encountered, as was discussed in
- chapter 7.

Parsing Rule applied: R-dass-1

```
******************************************************************************
ACTIVE MEMORY:

Words:                      sagen dass Agente
Syntactic Categories:       V     CL-V NG
Conceptualizations:         MTRO  NONO HUM1


MTRO =                                      NONO =
  CONCEPT MTRANS                              CONCEPT NONTHING
  ACTOR   HUM0 =
```

```
          CONCEPT   PERSON
          SPOKESMAN LOCO =
                      CONCEPT NATION
  TIME    INSO =
          CONCEPT INSTANCE
          DAY     TODAY

HUM1 =
  CONCEPT     PERSON
  NATIONALITY LOC1 =
              CONCEPT NATION
              #NAME   (iraq)
******************************************************************************

Parsing Rule applied: R-NEXT-WORD  (waehrend)
Parsing Rule applied: R-MAKE-SYM

******************************************************************************
ACTIVE MEMORY:

Words:                      sagen dass Agente waehrend
Syntactic Categories:       V     CL-V NG     PREP
Conceptualizations:         MTRO  NONO HUM1   DURO
******************************************************************************

Parsing Rule applied: R-NEXT-WORD  (eines)
Parsing Rule applied: R-MAKE-SYM

******************************************************************************
ACTIVE MEMORY:

Words:                      sagen dass Agente waehrend eines
Syntactic Categories:       V     CL-V NG     PREP     DET
Conceptualizations:         MTRO  NONO HUM1   DURO     NIL
******************************************************************************

Parsing Rule applied: R-NEXT-WORD  (Angriffes)
Parsing Rule applied: R-MORPH
Parsing Rule applied: R-MAKE-SYM

******************************************************************************
ACTIVE MEMORY:

Words:                      sagen dass Agente waehrend eines Angriffe
Syntactic Categories:       V     CL-V NG     PREP     DET   N
Conceptualizations:         MTRO  NONO HUM1   DURO     NIL   HARO
******************************************************************************

Parsing Rule applied: R-DET
```

```
*****************************************************************************
ACTIVE MEMORY:

Words:                   sagen dass Agente waehrend Angriffe
Syntactic Categories:    V     CL-V NG     PREP     NP
Conceptualizations:      MTRO  NONO HUM1   DURO     HARO
*****************************************************************************

Parsing Rule applied: R-NG

*****************************************************************************
ACTIVE MEMORY:

Words:                   sagen dass Agente waehrend Angriffe
Syntactic Categories:    V     CL-V NG     PREP     NG
Conceptualizations:      MTRO  NONO HUM1   DURO     HARO


MTRO =                                    NONO =
  CONCEPT MTRANS                            CONCEPT NONTHING
  ACTOR   HUMO =
            CONCEPT   PERSON
            SPOKESMAN LOCO =
                        CONCEPT NATION
  TIME    INSO =
            CONCEPT INSTANCE
            DAY     TODAY

HUM1 =                            DURO =              HARO =
  CONCEPT     PERSON                CONCEPT DURING      CONCEPT HARM
  NATIONALITY LOC1 =
                CONCEPT NATION
                #NAME   (iraq)
*****************************************************************************

Parsing Rule applied: R-PP

- The German Prepositional Phrase Rule, R-PP, leaves the NP in active
- memory, instead of the PREP, as in English and Spanish.  The "case"
- of the NP is marked, so that in this example, the "case" of "Angriffe"
- is "waehrund".  This case information is used to attach the PP.

*****************************************************************************
ACTIVE MEMORY:

Words:                   sagen dass Agente Angriffe
Syntactic Categories:    V     CL-V NG     NG
Conceptualizations:      MTRO  NONO HUM1   HARO
```

```
MTRO =                                        NONO =
  CONCEPT MTRANS                                CONCEPT NONTHING
  ACTOR   HUMO =
            CONCEPT   PERSON
            SPOKESMAN LOCO =
                        CONCEPT NATION
  TIME    INSO =
            CONCEPT INSTANCE
            DAY     TODAY

HUM1 =                               HARO =
  CONCEPT     PERSON                   CONCEPT HARM
  NATIONALITY LOC1 =
                CONCEPT NATION
                #NAME   (iraq)
********************************************************************************
```

Parsing Rule applied: R-AUX-NG

- The PP in which "Angriffe" appears is linked to "dass". Semantically, this means that the information DURING HARM is added to the representation stored under "dass".

```
********************************************************************************
ACTIVE MEMORY:

Words:                      sagen dass Angriffe
Syntactic Categories:       V     CL-V NG
Conceptualizations:         MTRO  NONO HARO
********************************************************************************

MTRO =                                        NONO =
  CONCEPT MTRANS                                CONCEPT NONTHING
  ACTOR   HUMO =                                DURING    HARO =
            CONCEPT   PERSON                                CONCEPT     HARM
            SPOKESMAN LOCO =
                        CONCEPT NATION
  TIME    INSO =
            CONCEPT INSTANCE
            DAY     TODAY

HUM1 =
  CONCEPT     PERSON
  NATIONALITY LOC1 =
                CONCEPT NATION
                #NAME   (iraq)
********************************************************************************
```

- At the end of the clause, the verb "toeteten" is encountered, and this verb is combined with the representation stored under "dass".

- In this sentence, the parser could not infer the action in the
- clause until the verb was read.

```
Parsing Rule applied: R-NEXT-WORD  (toeteten)
Parsing Rule applied: R-MORPH
Parsing Rule applied: R-MAKE-SYM

****************************************************************************
ACTIVE MEMORY:

Words:                    sagen dass Maenner toeten
Syntactic Categories:     V     CL-V NG      V
Conceptualizations:       MTRO  NONO HUM2    DEAO
****************************************************************************

Slot-filler Specialization Demon applied (because of RESULT
          filler of HAR1)
Prototype Failure Rule Applied:  S-DAMAGED-BY-PERSON (because attempted to
          fill RESULT-OF slot of DEAO with HUM1)
Parsing Rule applied: R-dass-2

****************************************************************************
ACTIVE MEMORY:

Words:                    toeten
Syntactic Categories:     V
Conceptualizations:       DEAO


DEAO =
  CONCEPT   DEAD
  RESULT-OF HAR1 =
            CONCEPT HARM-PERSON
            ACTOR   HUM1 =
                    CONCEPT     PERSON
                    NATIONALITY LOC1 =
                                  CONCEPT NATION
                                  #NAME   (iraq)
            OBJECT  HUM2 =
                    CONCEPT PERSON
                    GENDER  MALE
                    NUMBER  2
            RESULT  DEAO
  R1      HUM2
  DURING  HARO =
            CONCEPT    HARM
            SETTING-FOR DEAO
            PLACE      LOC2 =
                         CONCEPT LOCATION
                         NEAR    LOC4 =
```

```
                                     CONCEPT    LOCATION
                                     NATION-ADJ LOC3 =
                                             CONCEPT NATION
                                             #NAME   (iraq)
*******************************************************************************

- The parse continues until the last word, "nahmen" (seized).
- "Geisel" (hostages) is attached as the OBJECT of "seized", because
- semantics prefers to place the concept HOSTAGE into the OBJECT
- slot of GET-CONTROL, the representation of "nahmen".  Thus, since
- the case marking of "Geisel" is ambiguous between nominative and
- accusative, the parser chooses the accusative case, through semantic
- means.

*******************************************************************************

Parsing Rule applied: R-NEXT-WORD  (nahmen)
Parsing Rule applied: R-MAKE-SYM

*******************************************************************************
ACTIVE MEMORY:

Words:                     toeten und  Geisel nahmen
Syntactic Categories:      V      CONJ NG     V
Conceptualizations:        HAR1   NIL  HUM3   GET0
*******************************************************************************

Slot-filler Specialization Demon applied (because of OBJECT
        filler of GET0)
Parsing Rule applied: R-NG-V-CLASS

*******************************************************************************
ACTIVE MEMORY:

Words:                     toeten und  nahmen
Syntactic Categories:      V      CONJ V
Conceptualizations:        HAR1   NIL  GET0


HAR1 =
  CONCEPT HARM-PERSON
  ACTOR   HUM1 =
            CONCEPT     PERSON
            NATIONALITY LOC1 =
                          CONCEPT NATION
                          #NAME   (iraq)
  OBJECT  HUM2 =
            CONCEPT PERSON
            GENDER  MALE
            NUMBER  2
```

```
  RESULT  DEAO =
            CONCEPT   DEAD
            RESULT-OF HAR1
            R1        HUM2
            DURING    HARO =
                        CONCEPT     HARM
                        SETTING-FOR DEAO
                        PLACE       LOC2 =
                                      CONCEPT LOCATION
                                      NEAR    LOC4 =
                                                CONCEPT    LOCATION
                                                NATION-ADJ LOC3 =
                                                             CONCEPT NATION
                                                             #NAME   (iraq)

GETO =
  CONCEPT TAKE-HOSTAGES
  OBJECT  HUM3 =
            CONCEPT HOSTAGE
            NUMBER  SEVERAL
******************************************************************************

- The same conjunction rule as is used in English handles the conjunction
- in this sentence.  After "Geisel" is attached to "nahmen",
- the conjunction rule is applied, assigning the ACTOR of the killing
- to be the ACTOR of the action TAKE-HOSTAGES, and also marking
- the event TAKE-HOSTAGES as occuring DURING the raid (HARM).

Parsing Rule applied: R-CONJ

******************************************************************************
ACTIVE MEMORY:

Words:                  nahmen
Syntactic Categories:   V
Conceptualizations:     GETO
******************************************************************************

Parsing Rule applied: R-NEXT-WORD  (*PERIOD*)
Parsing Rule applied: R-MAKE-SYM

******************************************************************************
ACTIVE MEMORY:

Words:                  nahmen *PERIOD*
Syntactic Categories:   V      PUNC
Conceptualizations:     GETO   NIL
******************************************************************************

Parsing Rule applied: R-NO-PUNC
```

```
*******************************************************************************
ACTIVE MEMORY:

Words:                    nahmen
Syntactic Categories:     V
Conceptualizations:       GETO
*******************************************************************************


(Iran sagte heute dass irakische Agenten waehrend eines Angriffes in der
Naehe von der irakischen Grenze 2 Maenner toeteten und mehrere Geisel nahmen
*PERIOD*)
(nahmen)

Final representation:

GETO =
 CONCEPT TAKE-HOSTAGES
 OBJECT  HUM3 =
          CONCEPT HOSTAGE
          NUMBER  SEVERAL
 ACTOR   HUM1 =
          CONCEPT     PERSON
          NATIONALITY LOC1 =
                        CONCEPT NATION
                        #NAME   (iraq)
 DURING  HARO =
          CONCEPT     HARM
          PLACE       LOC2 =
                        CONCEPT LOCATION
                        NEAR    LOC4 =
                                  CONCEPT    LOCATION
                                  NATION-ADJ LOC3 =
                                               CONCEPT NATION
                                               #NAME   (iraq)
          SETTING-FOR GETO
HAR1 =
 CONCEPT HARM-PERSON
 ACTOR   HUM1
 OBJECT  HUM2 =
          CONCEPT PERSON
          GENDER  MALE
          NUMBER  2
 RESULT  DEAO =
          CONCEPT   DEAD
          R1        HUM2
          RESULT-OF HAR1
          DURING    HARO
MTRO =
```

```
 CONCEPT MTRANS
 ACTOR   HUMO =
           CONCEPT   PERSON
           SPOKESMAN LOCO =
                       CONCEPT NATION
 TIME    INSO =
           CONCEPT INSTANCE
           DAY     TODAY
 OBJECT  GETO

Total time:  124335 msecs.
NIL

Translation into English:

Iran said today that Iraqi agents killed 2 men. The agents seized a
number of hostages during a raid near the border with Iraq.
*
```

References

Akmajian, A., and Heny, F. *An Introduction to the Principles of Transformational Syntax*. Cambridge, Mass.: MIT Press, 1975.

The Automated Languages Processing Advisory Committe (ALPAC). *Languages and Machines: Computers in Translation and Linguistics*. Technical Report #1416, National Academy of Sciences, National Research Council, 1966.

Bar-Hillel, Y. The Present Status of Automatic Translation of Languages. *Advances in Computers*, 1960, **1**, 91-163.

Birnbaum, L., and Selfridge, M. *Problems in Conceptual Analysis of Natural Language*. Technical Report #168, Yale University Department of Computer Science, October 1979.

Bobrow, D.G., Kaplan, R.M., Kay, M., Norman, D.A., Thompson,H., and Winograd, T.. GUS, a Frame Driven Dialog System. *Artificial Intelligence*, 1977, 8(1), .

Bobrow, R.J., and Webber, B.L. Knowledge Representation for Syntactic/Semantic Processing. *Proceedings of the First Annual National Conference on Artificial Intelligence*, Stanford University, August, 1980.

Boitet, C., and Nedobejkine N. Recent Developments in Russian-French Machine Translation at Grenoble. *Linguistics*, 1981, (19), 199-271.

Bower, G.H., J.B. Black, and Turner, T.J. Scripts in memory for text. *Cognitive Psychology*, 1979, **11**, 177-220.

Carbonell, J. G., Cullingford, R. E., and Gershman, A. *Knowledge-Based Machine Translation*. Technical Report #146, Yale University Department of Computer Science, 1978.

Chandioux, J. METEO: An Operational System for the Translation of Public Weather Forecasts, pages 27-36. *FBIS Seminar on Machine Translation*, 1976. American Journal of Computational Linguistics, Microfiche 46.

Charniak, E. *Towards a Model of Children's Story Comprehension*. PhD thesis, Massachusetts Institute of Technology, 1972. Research Report #266.

Charniak, E. Ms. Malaprop: A Language Comprehension Program. *Proceedings of the Fifth International Joint Conference on Artificial Intelligence*, Cambridge, Mass., August, 1977.

Charniak, E. A Framed PAINTING: The Representation of a Common Sense Knowledge Fragment. *Cognitive Science*, 1977, **1**(4), 355-394.

Charniak, E. A Parser With Something For Everyone. *Proceedings of the Third Annual Conference of the Cognitive Science Society*, Berkeley, Calif., August, 1981.

Charniak, E. Context Recognition in Language Comprehension. In Lehnert, W., and

Ringle, M. (Ed.), *Strategies for Natural Language Processing*, chapter 10, pp 435-454, Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1982.

Chomsky, N. *Aspects of the Theory of Syntax*. Cambridge, Mass.: MIT Press, 1965.

Crain, S., and Steedman, M. On Not Being Led Up the Garden Path: The Use of Conntext by the Psychological Parser. In Dowty, D., Karttunen, L., and Zwicky, A. (Ed.), *(Title to be Announced)*, Cambridge: Cambridge University Press, 1982. To appear.

Cullingford, R. *Script Application: Computer Understanding of Newspaper Stories*. PhD thesis, Yale University, 1978. Research Report #116.

DeJong, G. *Skimming Stories in Real Time: An Experiment in Integrated Understanding*. PhD thesis, Yale University, May, 1979. Research Report #158.

Dyer, M. *In-depth Understanding: A Computer Model of Integrated Processing for Narrative Comprehension*. PhD thesis, Yale University, May, 1982. Research Report #219.

Gershman, A. *Knowledge-based Parsing*. PhD thesis, Yale University, 1979. Research Report #156.

Hemphill, L. *A Conceptual Approach to Automated Language Understanding and Belief Structures: With Disambiguation of the Word 'For'*. PhD thesis, Computer Science Department, Stanford University, 1975. Research Report #STAN-CS-75-534.

Hirst, G. *Semantic Interpretation Against Ambiguity*. PhD thesis, Brown University, December, 1983. Research Report #CS-83-25.

Katz, J., and Fodor, J. The Structure of a Semantic Theory. *Language*, 1963, **39**, 170-210.

King, M. Design Characteristics of a Machine Translation System.. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, Vancouver, B.C., Canada, August, 1981.

Lebowitz, M. *Generalization and Memory in an Integrated Understanding System*. PhD thesis, Yale University, October, 1980. Research Report #186.

MacDonald, R. R. *Machine Translation Research Project, General report 1952-1963*. Technical Report #30, Georgetown Univ. Machine Translation Res. Proj. Georgetown University, 1963.

Marcus, M. *A Theory of Syntactic Recognition for Natural Language*. PhD thesis, Massachusetts Institute of Technology, February, 1978.

Minsky, M. A framework for representing knowledge. In P. Winston (Ed.), *The Psychology of Computer Vision*, chapter 6, pp 211-277, New York: McGraw-Hill, 1975.

Riesbeck, C. Conceptual Analysis. In *Conceptual Information Processing*, Amsterdam:

North-Holland, 1975.

Riesbeck, C., and Schank, R.C. *Comprehension by Computer: Expectation-based Analysis of Sentences in Context*. Technical Report #78, Yale University Department of Computer Science, October 1976.

Schank, R.C. Conceptual Dependency: A Theory of Natural Language Understanding. *Cognitive Psychology*, 1972, 3(4), 552-631.

Schank, R.C. *Fundamental Studies in Computer Science*. Volume 3: *Conceptual Information Processing*. Amsterdam: North-Holland, 1975.

Schank, R.C. *Dynamic Memory*. Cambridge, England: Cambridge University Press, 1982.

Schank, R.C. and Abelson, R. *Scripts, Plans, Goals and Understanding*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1977.

Schank, R.C., and Birnbaum, L. *Memory, Meaning, and Syntax*. Technical Report #189, Yale University Department of Computer Science, 1980.

Schank, R., Birnbaum, L., and Mey, J. Integrating Semantics and Pragmatics. *Proceedings of the XIIIth International Congress of Linguists*, Tokyo, Japan, September, 1983.

Schank, R.C., Lebowitz, M., and Birnbaum, L. *Integrated Partial Parsing*. Technical Report #143, Yale University Department of Computer Science, 1978.

Slocum, J., and Bennett, W. *The LRC Machine Translation System: An Application of State-of-the-ARt Text and Natural Language Processing Techniques to the Translation of Technical Manuals*. Technical Report #LRC-82-1, Linguistics Research Center, University of Texas, July 1982.

Small, Steven. *Word Expert Parsing: A Theory of Distributed Word-based Natural Language Understanding*. PhD thesis, Department of Computer Science, University of Maryland, 1980.

Toma, P. SYSTRAN as a Multi-lingual Machine Translation System. In Commission of the European Communities (Ed.), *Overcoming the Language Barrier*, pp 569-581, Munchen: Vlg. Dokumentation, 1977.

Wilks, Y. An Artificial Intelligence Approach to Machine Translation. In Schank, R., and Colby, K. (Ed.), *Computer Models of Thought and Language*, chapter 3, pp 114-151, San Francisco: W.H. Freeman and Co., 1973.

Wilks, Y. A Preferential, Pattern-matching, Semantics for Natural Language Understanding. *Artificial Intelligence*, 1975, 6(1), .

Wilks, Y. *Seven Theses on Artificial Intelligence and Natural Language*. Technical Report #17, Instituto per gli Studi Semantici e Cognitivi, 1975.

Winograd, T. *Understanding Natural Language*. New York: Academic Press, 1972.

Woods, W. Transition Network Grammars for Natural Language Analysis. *Communications of the ACM*, 1970, **13**(10), .

Woods, W.A. Taxonomic Lattice Structures for Situation Recognition. *Theoretical Issues in Natural Language Processing - 2*, July, 1978.

Woods, W., Kaplan, R., and Nash-Webber, B. *The Lunar Sciences Natural Language Information System: Final Report*. Technical Report #2378, Bolt, Beranek and Newman, Inc., 1972.